WHAT'S UP WITH HIGH- AND LOW-PITCHED SOUNDS?

REFERENCE FRAMES USED IN THE CROSSMODAL CORRESPONDENCE BETWEEN AUDITORY PITCH AND VISUOSPATIAL HEIGHT

MICHAEL JAMES CARNEVALE

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

GRADUATE PROGRAM IN PSYCHOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

January 2015

© Michael James Carnevale, 2015

ABSTRACT

Low- and high-pitched sounds are perceptually associated with low and high visuospatial elevations, respectively. The spatial properties of this association are not well understood so two experiments were performed to investigate the following questions. Can low and high tones be used as spatial cues to upright for self-orientation? And what spatial frame(s) of reference is used to perceptually bind these crossmodal features? In experiment 1, participants' Perceptual Upright (PU) was measured with and without presented auditory orientation cues but there was no effect of sound. In experiment 2, the biasing effects of ascending and descending tones on ambiguous visual motion was measured when presented along both the gravitational and body reference frames, while participants sat either upright or laid on their side. There were effects of sound along both reference frames. A model predicting the axis of optimal association tentatively explains the findings of experiments 1 and 2.

ACKNOWLEDGMENTS

First, I would like to thank my family. Without their support and good faith I would not have the opportunities in life that I have. This thesis represents not only my dedication and hard work but their dedication, hard work, and love as well.

Second, I would like to thank Dr. Laurence Harris for giving me the opportunity to study in his laboratory and learn under his guidance. Thanks for being open minded with me and my projects and for helping me travel around the world to exotic locations including deserts, bicycle towns, and limestone alleyways.

Third, I would like to thank my labmates. We shared some great years of learning and great moments of delightful absurdity. I hope our comings and goings continue to crisscross and tie-up again in the future.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 General introduction

When one hears a bird's whistle through the rustling of trees, one might cheerily look up with a smile. When one hears the sound of a deep echoing thud, one might survey the horizon with their eyes for a heavy collision with the earth. Disregarding associations of meaning, the sound with the higher pitch originated above (i.e., the whistle) and the sound with the lower pitch came from below. As will be shown throughout this thesis, this is no accident and a longstanding literature demonstrates an intricate perceptual association across the senses between auditory pitch and visuospatial elevation. This thesis contributes to this literature by investigating two underlying issues. The first issue is: since high- and low-pitched sounds are respectively associated with high and low spatial elevations, can this association be used as a cue for spatial orientation? The second is the question of what frame of reference determines "high" and "low"? In other words, what spatial reference does the brain use to combine "high" and "low" sounds with "high" and "low" visuospatial elevations for multisensory integration?

First, I will introduce the concept of multisensory integration and its relation to audio-visual associations. Next, the concept of crossmodal correspondence will be defined, characterized, and divided into a typology. Then, the specific crossmodal correspondence between auditory pitch and visuospatial height will be reviewed. Finally, the spatial properties of the pitch-height correspondence will be discussed and the purpose and rationale for the experiments that comprise this thesis will be explained.

1.2 Multisensory integration and crossmodal correspondences

In everyday situations, our perceptual system is bombarded with information from the different sensory modalities. These sensory signals are processed into perceptual estimates composed of multiple features. For example, a sound can be interpreted perceptually in terms of features such as pitch, loudness, timbre, spatial location etc. whereas a visual stimulus can be interpreted in terms of colour, brightness, spatial frequency, shape, spatial location etc. This presents the brain with the problem of how and when to combine, or *bind*, these perceptual features into percepts of meaningful objects and events in the environment while keeping other perceptual information separate. The issue of how and when to combine perceptual features within a modality is referred to as the binding problem (e.g., combining the visual features of motion, shape, and color into a cohesive visual percept of a bouncing tennis ball; Triesman, 1980) while the issue of how and when to combine perceptual estimates and features across sensory modalities, the issue of interest to this paper, is referred to as the crossmodal binding problem (e.g., combining the visual percept of a tennis ball with the sound of it bouncing on the ground; Spence, 2011). The crossmodal binding problem is a central issue in the field of multisensory integration, which seeks to discover and understand how the senses interact in the brain.

The cognitive mechanisms of multisensory integration have been investigated and modeled both in terms of bottom-up and top-down processes. In terms of bottom-up processes, researchers have mainly focused on the temporal and spatial properties of multisensory integration. Typically, it has been found that signals from the different senses are more likely to be bound the closer to each other they are presented in time

(Jones & Jarick, 2006; Shore, Barnes, & Spence, 2006; van Wassenhove, Grant, & Poeppel, 2007). Multisensory integration has also been shown to occur under conditions of close spatial proximity (Bertelson, Vroomen, Wiegeraad, & de Gelder, 1994; Innes-Brown & Crewther, 2009; Jones & Jarick, 2006; Jones & Munhall, 1997). Other spatiotemporal features, such as the shared temporal structure of multimodal stimuli (i.e., that they have a perceived correlated pattern over the course of the event), have also been demonstrated to play a role (Radeau & Bertelson, 1987; Spence, 2007). Essentially, if stimuli appear at the same time, seem to originate from the same spatial location, and/or share spatiotemporal properties, it is more likely that they refer to the same object or event in the environment and are thus bound perceptually.

These spatiotemporal features of multisensory stimuli mentioned above are examples of redundant perceptual information and can be used advantageously by the brain. They are redundant in that the two senses, in reference to the same object or event, are providing estimates of the same type of information (e.g., the location, timing, number of objects or events) albeit with likely differences in precision. Since the same information is being gathered by multiple senses it is not specific to a particular modality and is thus referred to as an amodal stimulus property (Green & Angelaki, 2010; Bahrick, Lickliter, & Flom, 2004). The advantage of integrating multimodal redundant perceptual estimates of the same physical properties is that the multisensory representation can be more robust and precise compared to individual sensory estimates on their own (Ernst & Bülthoff, 2004; Trommershauser, Landy, & Kording, 2011). In fact, if the reliability of each perceptual estimate were measured, the increased accuracy and reliability of the perceptually integrated unified percept could be predictably modeled (Ernst & Banks,

2002). Conversely, non-redundant stimulus properties refer to stimulus properties that can only be uniquely conveyed by each sense (i.e., color can only be sensed through vision while pitch can only be sensed through audition) and are considered complementary. The advantage of binding non-redundant perceptual estimates is that it aids in providing a detailed cohesive representation of the environment with multiple layers of perception.

In conjunction with bottom-up processes governing multisensory integration, top-down processes have also been shown to play a role. One such mechanism is semantic congruency (Chen & Spence, 2010; Doehrmann & Naumer, 2008), a constraint that refers to whether or not percepts from different senses match in terms of their higher-level meaning or identity. To illustrate, multisensory integration between the image of a dog and the sound of a bark is more likely to occur than between the image of a dog and the sound of a cat's meow. Many of the studies on semantic congruency and multisensory integration have looked at ecological and naturalistic sounds (e.g., human voices) and how they bind with images and videos of related or unrelated content (Vatakis, Ghazanfar, & Spence, 2008; Vatakis & Spence, 2007, Chen & Spence, 2010). The concept of semantic congruency and its role in role in multisensory perception is discussed further in section 1.4.

Not all cases of multisensory integration can be explained entirely in terms of the bottom-up and top-down influences listed above however. Another factor playing a role in multisensory integration that has enjoyed a recent growth in research interest is *synaesthetic congruency*. Synaesthetic congruency refers to perceptual correspondences between elementary non-redundant features across different modalities. It has been

argued that multimodal stimuli that share synaesthetic congruency are more likely to be bound together, despite no obvious connection between their non-redundant stimulus properties (Spence, 2011). To illustrate, there is a perceptual compatibility between the auditory pitch of an object and its visual size such that larger visual objects are normally associated with lower-pitched sounds, and vice versa (Gallace & Spence, 2006). In the literature, this sort of perceptual compatibility effect is an example of what is referred to as a *crossmodal correspondence*. Crossmodal correspondences span a large range of sensory pairs such as vision and audition as mentioned above, vision and touch (e.g., Martino & Marks, 2000; Morgan, Goodson, & Jones, 1975; Simmer & Ludwig, 2009), audition and touch (e.g., Walker & Smith, 1985, Yau, Olenczak, Dammann, & Bensmaia, 2009), tastes/flavours and sounds (e.g., Bronner, 2011; Bronner, Bruhn, Hirt, & Piper, 2008; Crisinel & Spence, 2009; Mesz, Trevisan, & Sigman, 2011; Rudmin & Cappelli, 1983, Simmer, Cuskley, & Kirby, 2010), colour and smell (e.g., Gilbert et al., 1996; Kemp & Gilbert, 1997; Spence, 2010), or auditory pitch and smell (e.g., Belkin et al., 1997; Piesse, 1891; von Hornbostel, 1931). Some examples of crossmodal correspondences include associations between tactile size and auditory pitch (Walker & Smith, 1985), visual spatial frequency and auditory pitch (Evans & Treisman, 2010), sour and sweet tastes with high-pitched sounds (Crisinel & Spence, 2011), the colour green and the taste of "salt and vinegar" potato chips (Picqueras-Fiszman & Spence, 2011), or, perhaps my personal favorite, the smell of raspberries with the musical timbre of a grand piano (Crisinel & Spence, 2009). Crossmodal correspondences have been demonstrated to have varying effects on perception using a variety of experimental paradigms and examples. As will become clear throughout this thesis, since crossmodal correspondences

can be used to predict when multisensory integration is likely to occur, they play an important role in constraining the crossmodal binding problem (Spence, 2011).

1.3 Definition and properties of crossmodal correspondences

Crossmodal correspondences can be most generally defined as compatibility effects between attributes of stimuli from different sensory modalities that have some kind of predictive mapping between them. The general term includes instances of both redundant (e.g., sensing the location of an object with either vision or touch) and non-redundant (e.g., matching high-pitched auditory tones to small and/or bright visual objects) multimodal perceptual information but in this thesis it is used in its more colloquial sense of referring to non-redundant crossmodal feature level associations. What is unique about crossmodal correspondences is that the complementary features appear seemingly to be related arbitrarily and the mapping between them is uncertain. Unlike redundant cues, where perceptual estimates can be mapped from one sense to the other (e.g., seeing and feeling that there are three objects), mapping between complementary cues can only function in relative terms. To illustrate, low-pitched sounds are more readily associated with large rather than small visual objects, but the pitch does not actually provide the information necessary to determine in any absolute sense the actual size of the visual objects. Non-redundant crossmodal correspondences can be defined and categorized by the following three properties (terms from Spence, 2011; and Parise, 2012): polarity, relativity and universality.

For non-redundant crossmodal correspondences to work, the features in each modality must share the property of *polarity*. Polarity refers to the property that the value

of a perceptual feature can range across a perceptual dimension. For example, auditory pitch can range from lower-pitch to higher-pitch and the visual size of an object can range from very small to very large. This criterion is satisfied when the value of each feature can vary along polar dimensions and the dimensions of each perceptual feature can be mapped onto each other.

Related to polarity is the property of *relativity*, which highlights that the mapping between features in crossmodal correspondences is contextual. Relativity refers to the fact that the value of a particular stimulus along its polarized feature dimension is not absolute but exists in relation to other stimuli along the same dimension. To illustrate, a sound of a particular pitch can be defined as high pitched in relation to a sound of lower pitch but in contrast that same sound can be defined as low pitched in relation to a sound of higher pitch. The mapping between features across the senses depends on a perceptual context to give the associated features meaning.

The final property of complementary crossmodal correspondences is the property of *universality*. Universality refers to whether or not the crossmodal correspondence is universally perceived across most, if not all, observers. Cross-cultural studies have been done to investigate whether or not certain crossmodal correspondences hold up across different cultures to determine if they are universal. Much of this work has investigated the universality of crossmodal correspondences between the sound of certain spoken words and meaningless visual stimuli, a phenomenon known as *sound symbolism*. Research on sound symbolism actually goes back to the origins of study into the topic of crossmodal correspondences in general and dates back to work by Kohler in 1929 where he found that when presented with a rounded amoeba-like visual object and an angular

object with pointed edges, subjects were more likely to pair the round globular object with the nonsense word “Maluma” than with the nonsense word “Takete”, rather than vice versa (Figure 1). Subjects report that there is something intuitively related between the features of the visual object and the sound of the two names. Recent cross-cultural research suggests that non-Western and isolated cultures (Bremner et al., 2013) also make this phonetic and visual pairing. Other studies, conducted with slight variations of these stimuli, suggest that the crossmodal correspondence of sound symbolism is universally perceived (Davis, 1961; Gebels, 1969; Hiton, Nichols, & Ohala, 1994; Osgood, 1960; Rogers & Ross, 1968; Taylor & Taylor, 1962).

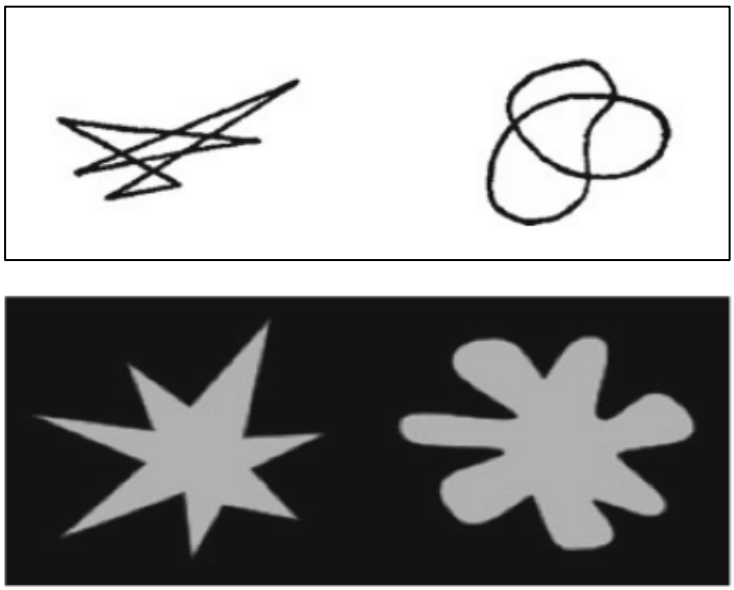

Figure 1. Top: Visual stimuli used in Kohler (1929). Participants when asked were more likely to label the left object as “takete” and the right object as “maluma”. Bottom: Visual stimuli used in Ramachandran (2001). Participants almost unanimously labelled the left object “kiki” and the right object “bouba”.

1.4 Types of polar crossmodal correspondences

From the research accumulated across studies on crossmodal correspondences it appears that their perceptual origins may have different underlying principles and it has recently become a matter of interest to categorize them (Spence, 2011). These underlying differences potentially reflect different neural substrates and may have qualitatively different effects on human perception (Westbury, 2005). The following three types of crossmodal correspondences have been proposed: structural correspondences, statistical correspondences, and semantically mediated correspondences.

Structural correspondences: Structural correspondences refer to crossmodal correspondences that arise from the structural characteristics of the neural system used to code the sensory information. These neural substrates are suggested to reflect intrinsic properties of the perceptual system's organization (Marks, 1978). A clear example of the concept of structural correspondences comes initially from S.S. Stevens (1957) who discussed the fact that the firing rate of neurons increases as a positive function of stimulus intensity regardless of the modality. Stevens suggested that this might relate to an underlying neural connection that could lead to crossmodal correspondences such as that between loudness and brightness (Marks, 1987).

Another potential example of the neural underpinnings of structural correspondences comes from the behaviour of the inferior parietal cortex. This brain region appears to play a role in general purpose coding of the magnitude of various perceptual features regardless of which sensory modality provides the input (Walsh, 2003). The inferior parietal cortex appears to code the magnitude of various spatial, temporal, and qualitative features of sensory information in a common metric and it is possible that this structure underpins certain polar crossmodal correspondences.

A final example of the possible underpinnings of crossmodal correspondences based on basic structural characteristics of the nervous system is the proximity or interconnectedness of brain areas. Nearby brain areas or areas with reciprocal connections may be responsible for some crossmodal correspondences (Ramachandran & Hubbard, 2001; Rouw & Scholte, 2007). However, crossmodal correspondences that arise from the connectedness and proximity of brain areas do not lend any predictive power as to which crossmodal polar features will be congruent or incongruent (i.e., should high or low pitch be paired with large or small visual size?).

Statistical correspondences: Statistical correspondences are essentially crossmodal correspondences that result from perceptual learning. Some multisensory stimuli have features that are often experienced in a correlated fashion as a result of the physical features in the environment being correlated through physical laws. For example, with the crossmodal correspondence between auditory pitch and visual size, larger objects in fact typically *do* resonate with lower frequencies and vice versa. Through consistent experience, the perceptual system adapts to such correlations and learns to expect that the lower-pitched sound should be paired with the larger object. This process of multisensory perceptual learning has been observed with a variety of stimuli (Adams, Graf, & Ernst, 2004; Stocker & Simoncelli, 2006; Weiss, Simoncelli, & Adelson, 2002) and it appears that there are many crossmodal correspondences potentially based on such learned associations (Bernstein & Edelstein, 1971; Gallace & Spence, 2006; Marks, 1987a, 1987b, 1989, 2004; Melara & O'Brien, 1987). This is an adaptive response as the system learns to better predict the features of the external environment.

Considering that statistical correspondences reflect the properties of the environment, these crossmodal correspondences should be universal. There are, however, examples of artificially induced crossmodal correspondences produced in laboratory settings using specified training regimes (Baier, Kleinschmidt, & Muller, 2006; Ernst, 2007; Zangenehpour & Zatorre, 2010). For example, Ernst (2007) entrained a perceptual association between tactile stiffness (i.e., the physical resistance a robotic arm provided in response to pressure put forth by the participant) and visual brightness, such that increased tactile stiffness became arbitrarily associated with increased visual brightness. Such artificial crossmodal correspondences, induced using repetitive paired presentations of stimuli (in as little training time as 45 minutes), are not "universal", and it is often found that they revert back to their initial associations (or no association at all) after some time post-training. Thus, it appears possible to both train new crossmodal correspondences (i.e., where the perceptual features had no previous association) and to re-train previously established correspondences.

It has become popular to model the process and properties of statistical correspondences in terms of Bayesian integration theory which posits that the brain combines stimuli in a statistically optimal manner by combining prior perceptual knowledge and/or expectations (referred to simply as 'priors') with sensory information and weighting each of them in proportion to their reliabilities (Ernst & Bülthoff, 2004; Ernst, 2006, 2007). In relation to crossmodal correspondences, these priors represent the sensory system's prior expectations that certain crossmodal stimuli "go together".

In terms of the Bayesian view, the mapping between sensory signals can be modeled with a coupling prior, which represents the expected joint distribution of the

signals. Figure 2, which is adapted from Ernst (2007), shows a 2-D representation (where the horizontal-axis refers to the visual sensory component S_v and the vertical-axis refers to the auditory sensory component S_a of the true physical stimulus denoted X) of the perceptual process of integrating bimodal sensory estimates using a coupling prior. In this example, a sensory stimulus having both a visual and an auditory property (represented in the top row as likelihood distributions where the standard deviation of S_v is double that of S_a) interacts with three possible coupling priors (the middle row shows three possible coupling priors by column) to yield one of three final percepts (bottom row). In the case of bimodal integration described here, the coupling prior distribution can be represented as a 2-D Gaussian distribution with infinite variance along the positive diagonal (the identity line where visual and auditory components are equal). Along the negative diagonal the variance can vary, representing the degree to which the bimodal cues are perceptually integrated (where the strength of coupling increases as this variance decreases). In this example, the prior on the left has an infinite variance along the negative diagonal, thus the bimodal stimuli are treated as independent with no interaction between them. With the prior on the right column, the variance is 0 along the negative diagonal and the bimodal cues are completely fused into one integrated multisensory percept. The middle column prior represents an intermediate coupling prior value, which leads to a coupling interaction of the two without entirely collapsing the two unimodal signals into one entirely integrated sensory fusion.

Crossmodal correspondences can be thought of in this framework in the sense that there are coupling priors that determine how much particular multisensory signals should be integrated. As explained in the previous paragraph, depending on the strength of the

coupling prior, some multisensory stimuli are integrated into unified percepts while others will show partial integration, or none at all. Thus, crossmodally correspondent stimulus pairs could be said to have a stronger coupling prior than non-correspondent crossmodal pairs.

There are cases of crossmodal correspondence that researchers suggest may be programmed innately as association priors in the brain but this hypothesis is as yet unresolved. For example, psychophysical studies involving 3-4 month olds have demonstrated preferential looking at visual stimuli (varying in visuospatial height and visual sharpness) when they were paired with an auditory stimulus of congruent pitch, suggesting that these correspondences may be innate (Walker, 2010). Other developmental studies have looked at crossmodal correspondences early in life including work by Lewkowicz and Turkewitz (1980) investigating the loudness-brightness correspondence in 20 day-old infants, and work by Mondloch and Maurer (2004) where they looked at auditory pitch and visual size and brightness. The problem with suggesting that these priors are innate however is that training these correspondences can happen in as little as 45 minutes and the youngest infants tested were 20 day-old infants The jury is therefore still out on whether or not these infants were born with these priors.

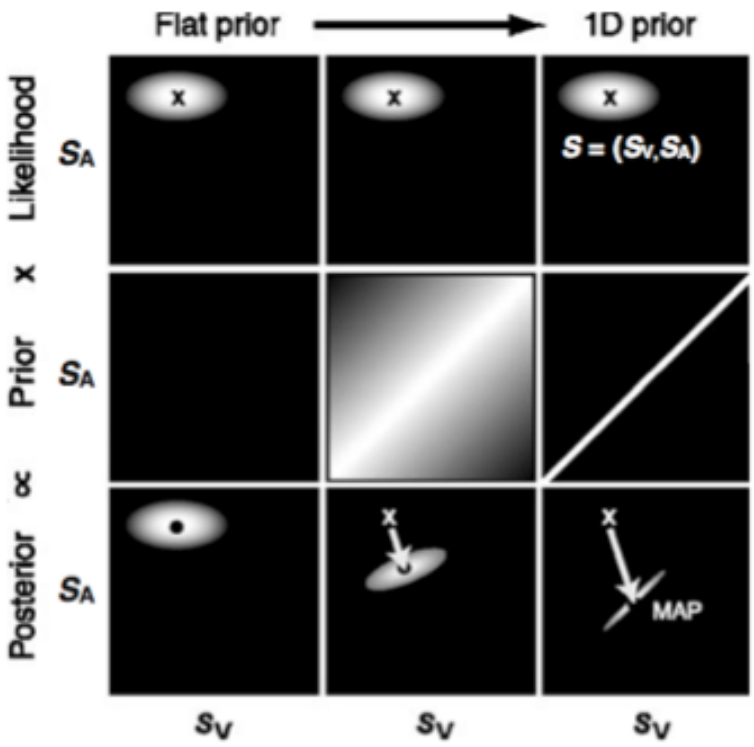

Figure 2. Schematic illustrating how bimodal sensory information interacts with coupling priors to yield final percepts (see section 1.4 *statistical correspondences* for full description). The horizontal axis represents the visual sensory component S_v and the vertical axis represents the auditory sensory component S_a of the true physical stimulus denoted X, or S=(S_v, S_a). Top Row: Sensory estimate for bimodal stimulus X. Middle Row: Coupling priors with infinite variance on the negative diagonal (left), intermediate variance (middle), and zero variance (right), where as the variance decreases the strength of coupling increases leading to a higher degree of multisensory integration. Bottom Row: Final multisensory percepts after influence of the associated coupling prior. Bimodal estimates remain entirely separate (left), bimodal estimates are somewhat integrated (middle), and bimodal estimates are integrated fully into one unified multisensory percept (right). Figure adapted from Ernst et al. (2007).

Semantic Correspondences: Semantic correspondences refer to those crossmodal correspondences that seem to originate from associations found within the spoken language of the user (Long, 1977). Probably the clearest example of a crossmodal correspondence that has a semantic correspondence is that between auditory pitch and visuospatial height (Eitan &Timmers, 2010; Stumpf, 1883). The semantic relation is the fact that reportedly in many languages people refer to auditory pitch as "high" or "low" to denote high and low frequency sounds, despite the fact that "high and "low" are linguistic prepositions that describe physical location, while the pitch of a sound does not inherently have any spatial properties at all. Such quirks of language are suggested to lead to a semantic association in the brain that may relate the features of two senses and lead to a crossmodal correspondence. Conversely, considering that many languages across cultures have similar such quirks of language it can be argued that perhaps the lexicon itself actually reflects a crossmodal correspondence acquired through other means such as statistical co-occurrence. Perhaps both influences play a role and the correspondence is supported both in language and through other means simultaneously.

Crossmodal correspondences need not necessarily be compartmentalized exclusively as arising from any one of the above mechanisms but most likely have overlapping origins. The crossmodal correspondence between auditory pitch and visuospatial height is thus an interesting example, as current research seems to suggest that it exemplifies properties of all three types of crossmodal correspondence described above.

Having established some background material on crossmodal correspondences in general, I now shift focus to the crossmodal correspondence of particular interest to this

thesis, the crossmodal correspondence between auditory pitch and visuospatial height.

1.5 The crossmodal correspondence of auditory pitch and visuospatial height

The association between auditory pitch and visuospatial height is one of the most well-known and perceptually robust crossmodal correspondences in the literature (Spence, 2011). Spanning a history of published research of over a century this crossmodal correspondence has become a universal hallmark of human cognition. As will be shown over the course of this review, the spatial connotation of auditory pitch whereby higher-pitched sounds are associated with higher spatial elevations, and vice-versa, meets all the aforementioned listed characteristics of crossmodal correspondences (i.e., polarity, relativity, and universality) and interestingly appears to originate from all three types of crossmodal correspondence mechanisms (i.e., structural, statistical, and semantic).

The earliest work in the literature on the association between auditory pitch and spatial elevation can be traced to the cross-cultural study of language; far before the term "crossmodal correspondence" was coined. As mentioned in the previous section, Stumpf (1883) pointed to the fact that this association holds up across languages and cultural forms. The musical staff for example, places "high notes" graphically higher in the scale than "lower notes". Singers and musicians "reach" for higher notes as if reaching towards the sky. One can even observe this metaphor when a young choir student stands on their tiptoes to sing the highest note of a musical phrase. Stumpf (1883) noted that auditory pitch does not have any intrinsic spatial properties and acknowledged how peculiar the universality of this association is. He suggested that there must be some explanation

outside of language to explain this seeming coincidence.

Stumpf's (1883) work inspired psychophysicists who further investigated the perceptual effects of this association. In a highly influential paper, Pratt (1930) found that subjects' auditory localization in the vertical plane was highly biased depending on the frequency content of presented sounds. Pratt also noted that human auditory localization in the vertical plane is much less accurate than in the horizontal plane. This is due to the fact that the observer cannot rely on binaural cues such as timing and intensity differences to localize a sound in the vertical plane. Instead they must rely on the less reliable spatial auditory cues generated by the sound-filtering properties of the external ear (one component of the spatial sound filtering cues referred to as head-related-transfer functions; Blauert, 1997). Pratt had subjects localize the position of tones sounded by a hidden loudspeaker having variable location (participants were informed of the location of the different speakers used) using a numbered scale ranging from the floor to the ceiling. Five tones were used with frequencies of 256, 512, 1024, 2048 and 4096 Hz and he found that subjects consistently localized them in that order from bottom to top. To avoid confounds related to using the number line, Trimble (1934) replicated this effect by asking subjects to verbally report the vertical displacement of presented tones (which came from the same veiled loudspeaker source) and to draw on a chart the apparent course of "ascending" and "descending" presented tones. As expected, higher-pitched tones were localized higher in space, and "ascending" tones were drawn with an upward trajectory, and vice-versa. In an attempt to avoid any confounding effects of instructions, Mudd (1963), presented subjects with a pair of tones of different frequency and simply asked them to move a peg (representing the position of the first tone) from a starting

location to another, to represent the spatial displacement in subsequent tones. He found that participants placed the peg up and slightly to the right when presented with a higher-pitched second tone and down and to the left for a lower-pitched second tone. Later, Roffler and Butler (1968) replicated again Pratt's initial finding with nine tones (250, 400, 600, 1400, 2000, 3200, 4800, 7200 Hz), giving participants discreet localization options, but varied participants' distance from the panel and also manipulated viewer orientation (i.e., participants were positioned upright, supine, or laying on their right side). They found the pattern persisted despite these manipulations and found errors in localization primarily along the body axis (the role of body orientation on the pitch-height correspondences will be expanded upon in section 1.5). Furthermore, they tested congenitally blind subjects and 4-5 year old children and found similar results.

Following the auditory localization studies inspired by Pratt (1930), novel research methods were used to study the cognitive effects of the association between auditory pitch and spatial elevation, which incidentally played an influential role in facilitating research on crossmodal correspondences in general. A seminal study by Bernstein and Edelstein (1971) found, using a speeded classification task, that participants responded slower to visual targets when their spatial elevation was incompatible with the pitch of an auditory tone. In their task, participants had to indicate whether a visual target was flashed below or above a fixation cross while a task-irrelevant auditory tone (either 100 or 1000 Hz) was simultaneously presented. The task-irrelevant sound either sped up or slowed participants' speed of classification, relative to a no-sound block, depending on whether the sound was congruent or incongruent with the spatial elevation of the target. This effect of task-irrelevant sounds facilitating or detracting from

the speeded classification of visual targets that vary in spatial elevation has been replicated in a number of studies since (see Melara & O'Brien, 1987; Ben-Artzi & Marks, 1995; Patching & Quinlan, 2002; and Evans & Treisman, 2010). Other researchers who noted perceptual links between other crossmodal stimulus pairs used the speeded classification task and found similar effects (see Marks, 2004, for a review of crossmodal correspondence research utilizing the speeded classification paradigm). Basically, participants' speeded classification performance was hindered when the associative feature of targets (usually visual) was incongruent with the related crossmodal feature of a task-irrelevant stimulus (usually auditory).

Other studies on the pitch-height association have shown the cognitive effects of semantic and lexical stimuli using similar methods. For example, the visual speeded classification effect described in the previous paragraph has been shown to persist even when participants were presented with the spoken words "High" and "Low" (Gallace & Spence, 2006). This semantic link was also demonstrated in a STROOP task where participants had to report if they saw the words "UP", "TOP", "DOWN", or "BOTTOM" while presented with a task-irrelevant high or low tone (Melara & Marks, 1990). This semantic effect however does not appear to influence perception by the same mechanism as auditory pure tones (e.g., Maeda et al., 2004) as neuroimaging research shows that these effects act at different processing levels (i.e., true perceptual bias vs. top-down semantic level bias) to influence subjects' behaviour during different experimental tasks (Sadaghiani et al., 2009). Their fMRI study (Sadaghiani et al., 2009) revealed feature-level audiovisual pitch-height interactions in left human motion complex (hMT+/V5+) whereas speech stimuli activated the right intraparietal sulcus. There is currently

consensus in the literature that true multisensory integration can only be said to occur when there are true feature-level interactions of stimuli (Spence, 2010). The heterogeneity in the underlying mechanisms governing different examples and types of crossmodal correspondences therefore suggests that they can operate at different processing levels which may potentially interact, and that all crossmodal correspondences are best understood on a case-by-case basis.

A few theories have been put forward to explain the reaction time effects found in relation to the speeded classification task. In all the speeded classification tasks, participants are expected to respond as quickly as possible to one stimulus while ignoring a supposedly task-irrelevant stimulus. If the feature dimensions of each of the multimodal stimuli were processed independently, the response times to the target stimuli should not be affected by variations in the feature dimension of the crossmodal stimulus (Garner, 1974). Martino and Marks (1999) characterized these dimensional interactions in terms of information accrual. They suggest that to make a response, a certain criterion threshold must be achieved. When congruent information is received from both channels, the information interacts and the speed of reaching the response threshold is speeded up. Marks (2004) characterized these interactions in terms of attentional resources and suggested that the incongruent crossmodal stimulus feature, which does not "fit", diverges attention and interrupts participants' timed responses. Miller (1991) also alluded to the idea of statistical facilitation (put forth by Raab, 1962), which states that when there are two redundant sources of information, one of them will be processed faster than the other, leading to an overall statistical likelihood of faster processing.

Researchers have also demonstrated that the perceptual effects related to the

pitch-height association extend to motion, where ascending tones are associated with upward visual motion and descending tones are associated with downward visual motion. In a well-known study by Maeda et al. (2004), participants judged the direction of visual motion (upward vs. downward) which was composed of two superimposed, oppositely moving sinusoid gratings, accompanied by a tone that was either ascending or descending in frequency, or broad-band noise. The two superimposed visual gratings varied in contrast ratio, producing visual stimuli that ranged from clearly completely downward to completely upward visual motion, including a range of mixed ambiguous motion stimuli. Gratings with ambiguous motion presented with an ascending pitch were more likely to be perceived as upward motion, and those accompanied by descending pitch as downward motion, whereas noise caused no directional bias. This effect has also been found with apparent motion (i.e., the illusory perception of motion induced when a target flashes in one location and then immediately flashes in a nearby location) where ascending tones bias visual motion processing upwards and vice-versa (Sadaghiaini et al., 2009).

Since the pitch-height crossmodal correspondence is so widespread, robust, and affects the observer in a range of perceptual and cognitive tasks, it begs the question, is this perceptual association innate? Walker et al. (2009) demonstrated that 3-4 month old infants preferentially looked at visual stimuli, varying in visuospatial height and visual sharpness, when they were paired with an auditory stimulus of congruent pitch, and concluded that this correspondence is innate. As stated at the end of the previous section however, since it has been demonstrated that crossmodal correspondences and perceptual associations can be trained in as little as 45 minutes, this association found in infants may

actually reflect an adaptation to the statistical properties of the environment. To investigate this possibility, Parise et al. (2014) recorded sounds from the natural environment by having a participant walk around various natural (e.g., park) and unnatural (e.g., city) landscapes with directional microphones attached to their heads: one pointing up and the other pointing down. They indeed found that statistically sounds with higher-frequency content tend to come from above and sounds with more low-frequency content tend to come from below, suggesting that the pitch-height association is a statistical correspondence. They posited that the ground might act as a band-pass filter, absorbing the high frequencies and reflecting or reverberating the low frequencies. They also noted that the anatomical frequency filtering properties of the external ear, and the head-related transfer functions that they produce by modifying the spectra of the sounds (Batteau, 1967), accentuate this already existent frequency-elevation relationship found in real-world sounds. Parise et al. (2014) analyzed a set of forty-five head-related transfer functions recorded using in-ear microphones (taken from recordings from an audio database, Algazi et al., 2001) and confirmed that identical sounds presented from higher in space were associated with more energy in the higher-frequency range, and vice-versa for sounds coming from lower in space (Figure 3). In the conclusion of their paper they posit that these filtering properties of the ear may have in fact evolved to highlight the relationship between auditory frequency and spatial elevation in the environment and make it a more useful spatial cue. This is not entirely impossible, as it has been found that the filtering properties of the eye also exaggerate the statistical properties of the visual scene and consequently enhance perception (Burge & Geisler, 2011). In conclusion, the auditory pitch and visuospatial height correspondence appears to be a statistical

correspondence (in which the correspondence is reflected in the statistical properties of the natural environment), a structural correspondence (in the loose sense that the external ear filters sounds to emphasize these statistical likelihoods), and a semantic correspondence as demonstrated in this section.

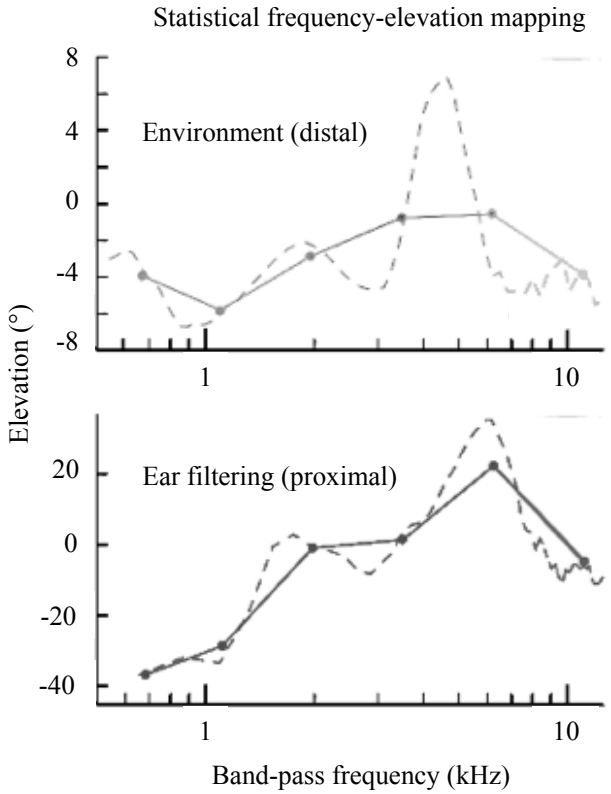

Figure 3. Statistical mapping between sound frequency and spatial elevation (relative to the head). Top: Sound spectra recorded in the environment using directional microphones attached to a participant's head. Bottom: Sound spectra recorded using in-ear microphones, based on head-related transfer functions taken from an audio database (Algazi, et a., 2001). The solid lines represent binned data by band-pass frequency while the dashed lines represent mathematical fits to the non-binned data. Figure adapted from Parise et al. (2014).

1.5 Spatial properties of the auditory pitch-visuospatial height correspondence

While this review of an extensive literature has shown that there are significant perceptual and cognitive consequences of the pitch-height crossmodal correspondence, there are still unanswered questions regarding its underlying perceptual mechanisms. Of interest to this thesis is the fact that throughout the literature limited work has been done to investigate the underlying spatial mechanisms, despite the fact that it is an inherently spatial association. This may reflect the fact that until recently the formal concept of a crossmodal correspondence was not well researched as a topic in and of itself but rather comprised numerous disjointed studies on such peculiar perceptual associations. Now that there is some consensus that crossmodal correspondences truly play a role in multisensory integration and are not simply quirks of language or the human cognitive system (Spence, 2011), there can be a more detailed look at their underlying perceptual mechanisms, further implications, and even practical applications. This section highlights some findings concerning the spatial properties of the pitch-height association including recent influential work and leads to the next sections which outline the purpose of the two studies presented in this thesis.

What is the spatial representation of tones in the brain? As Pratt (1930) and subsequent researchers have shown, tones tend to be visuospatially localized as a function of their pitch. Rusconi et al. (2006) elaborated on this and proposed that tones are represented in the brain along a "mental tonal axis", an analogue to the previously reported "number line" (Dahaene et al., 1993). In their study they had participants perform a timed pitch discrimination task where the independent variable was the mapping of response keys. High and low tones were either associated with a

visuospatially high or low response key respectively, (buttons “q” and “spacebar” respectively on a keyboard that was placed flat on a table in front of participants) or vice-versa, where high and low tones were associated with visuospatially low or high response keys, respectively. Predictably, they found that participants responded more quickly when the response key and tone corresponded spatially. They reasoned that the spatial cognitive representation of tones matched with the spatial layout of the response keys and facilitated faster responses.

In the spirit of Pratt’s (1930) work on tonal localization, there is some research that has looked at how participants localize auditory tones when the participant’s orientation relative to the external world is manipulated. By manipulating participants’ orientation relative to the environment researchers can set the frames of reference in which perceptual cues are encoded to determine what constitutes “upright” in spatial conflict (Aubert, 1861; Howard, 1982). When one is tilted over, “upright” may be defined relative to one’s own body (i.e., towards the head is “high” and towards the feet is “low”), or, in relation to gravity (i.e., the direction an object falls under the pull of gravity. The head is the initial reference frame for sounds because anatomically the ears are attached to head, but this does not mean they are necessarily coded craniotopically. By manipulating participants’ orientation, researchers are able to ask the fundamental question of what representation of “high” and “low” is used when localizing tones of different frequencies. With this logic, Roffler and Butler (1968) extended Pratt’s work and had participants lie on their right side and localize tones onto a modified visual panel (Figure 4). They found that participants mislocalized the high tones with respect to both gravitational and head-centric “up”, although the latter was more prominent.

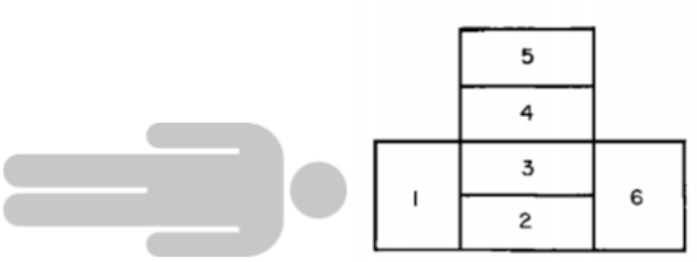

Figure 4. Visual panel used for auditory localization in Roffler and Butler (1968). The human silhouette shows that participants laid on their right sides during this localization task. During the actual experiment participants' heads were rested in front of the panel between sections 2 and 3. For high tones, participants were more likely to report that they perceived their origin as coming from sections 4, 5, and 6. There was a greater preference however for section 6, suggesting that the body reference frame was the dominant reference for "up" in this task. Figure adapted from Roffler and Butler (1968).

Very recently Parise et al. (2014) also looked at auditory localization under different spatial orientations (upright, tilted 45°, tilted 90°) and employed much more contemporary methods and theory than Roffler and Butler (1968). In their task, participants had to localize bursts of white noise and band-pass filtered noise (<0.8, 0.8-1.4, 1.4-2.5, 2.5-4.5, 4.5-8, >8 kHz) presented via a 4x4 speaker grid occluded by a sound-transparent black foreground. Unlike previous studies (e.g., Pratt, 1930; Roffler & Butler, 1968), their localization task afforded greater freedom of localization response, where participants used a laser to point to where they perceived the sound to originate. Predictably, given the literature reviewed above, when upright, participants mislocalized sounds as a function of their frequency energy, and confirmed the *frequency elevation*

mapping, as they coined it. What is interesting is participants' localization behaviour when their bodies' orientation was tilted from gravitational upright. When tilted sideways, participants mislocalized sounds in a way consistent with both being "high" and "low" relative to the head *and* relative to gravity, suggesting that both frames of reference played a role in their frequency elevation mapping for localization. Parise et al. (2014) explained this as being the result of combining distal (i.e., with respect to the external environment, here represented by the axis of gravity) and proximal (i.e., with respect to the self, here represented by the head axis) reference frame priors used for relating auditory pitch to visuospatial height (Figure 5A, and see section 1.4 for an explanation of the concept of priors). Parise et al. (2014) suggest that each localization prior has its own frequency elevation mapping with respect to its reference frame and it appears, based on participants' localization behaviour, that when they are not spatially congruent (i.e., when the gravity and head reference frames are not aligned), they are combined in a weighted fashion to produce a singular localization of sound origin in between the two "ups". Presumably these reference frame priors are always weighted this way, but it is only observable when they are misaligned. While this theory appears satisfactory in explaining auditory localization behaviour, it remains unclear which reference frames are used when combining auditory tones with distinct visual stimuli.

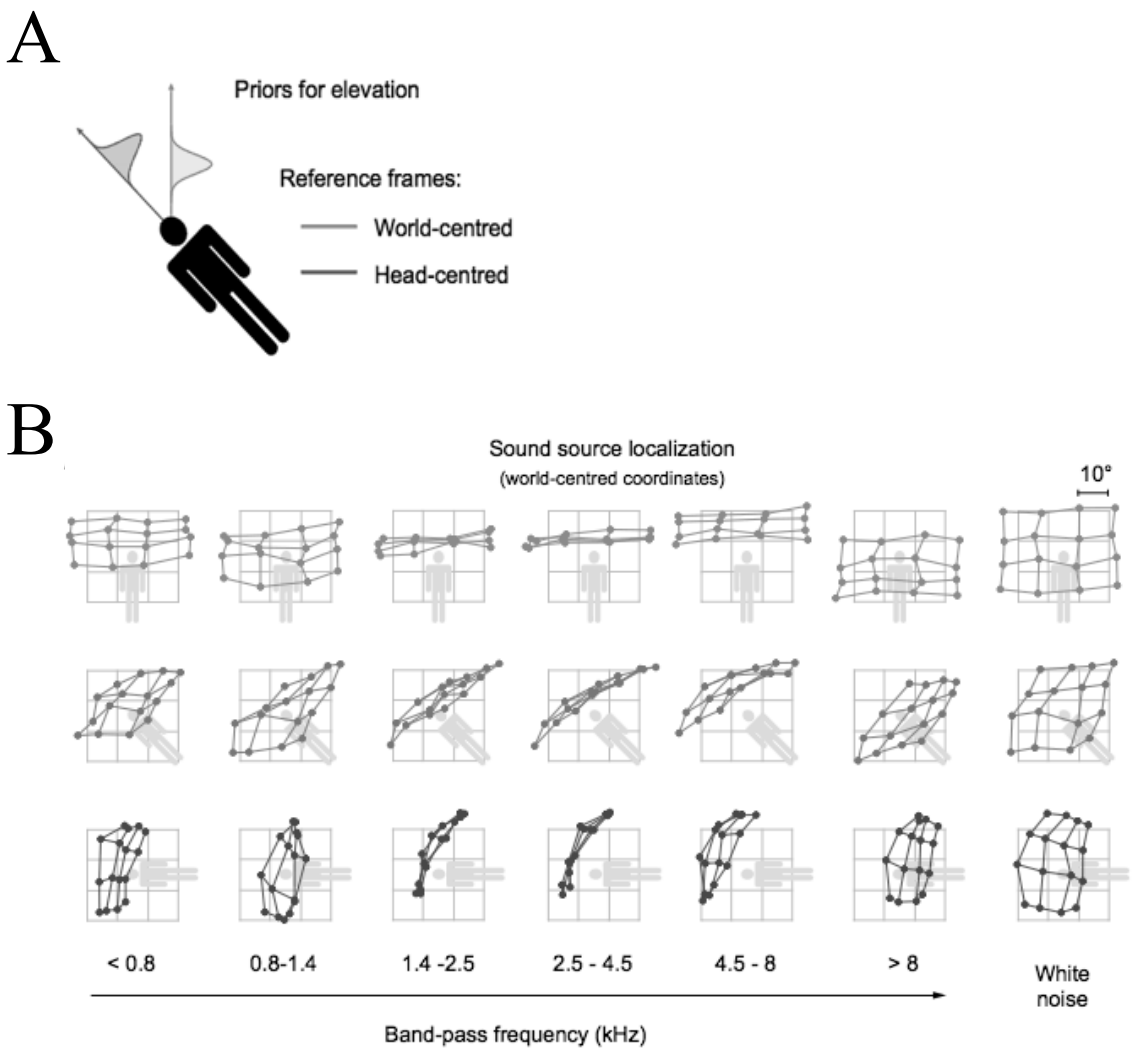

Figure 5. (A) World and head centered reference frames and their associated frequency by elevation localization priors. For each reference frame the priors are represented with Gaussian distributions, whose mean represents the expected elevation given the frequency of the sounds. (B) Localization behaviour of participants for different sounds by frequency where the grid shows the 4x4 speaker grid and the meshed points show participants' average responses for those speakers. Participants' responses suggest that auditory localization depends on a weighted combination of both the gravitational and head centered reference frames. Both figures adapted from Parise et al. (2014).

Maeda et al. (2004) investigated the spatial properties of the pitch-height correspondence in the visual motion domain, varying not the orientation of the participant but rather the orientation of the visual stimulus. In Maeda et al.'s (2004) work where they presented a motion discrimination experiment with ambiguous oppositely moving visual gratings, they manipulated the orientation (horizontal, -45°, +45° upright) of the visual presentation in a separate experiment (see section 1.4 for more detail on Maeda et al., 2004). They found that the perceptual biasing effect of ascending and descending tones gradually degraded as the visual orientation of the gratings shifted further from upright. This suggests that as the reference frames used for binding pitch to height (i.e., gravitational and body-centered) and the visual axes of visual motion became spatially misaligned multisensory integration gradually degraded. In this experiment however, where participants sat upright, it remains unclear which reference frame was used to define the tonal axis for integrating the auditory stimulus with visual motion perception.

This thesis focuses on two questions related to the spatial properties of the pitch-height correspondence that are not well understood. The first question, which is entirely novel to this thesis is, since auditory tones are associated with high and low spatial elevations (and as mentioned, this is supported by the statistical properties of the environment and perhaps even structural components of the human organism), can high and low tones be used by the brain as a reference cue to determine one's perceived direction of "upright"? The second question is, when the brain integrates auditory tones with "high" and "low" spatial elevations for visual motion perception, what frames of reference are used in determining the spatial axis of "high" and "low"?

1.6 Purposes and Rationale

1.6.1 Experiment 1 – Can low and high tones be used as auditory cues to self-orientation and influence the perceptual upright?

In experiment 1, the potential for the pitch and spatial elevation crossmodal correspondence to play a role in spatial orientation is investigated. The rationale for this study is that since the pitch-height correspondence appears to influence perception (e.g., the vertical localization of sounds in the environment) and cognition in a range of tasks and appears to arise from adapting to the statistical properties of the environment (where high tones indeed tend to come from higher in space and vice-versa, Parise et al., 2014, see section 1.4), it is possible that sounds of different frequency can provide the brain with a spatial reference to high and low space, or "up" and "down". This spatial reference could then potentially be used as a cue for self-orientation, a factor that is known to play a role in organizing perception (see below in this section). Thus, measuring and comparing participant behaviour that depends on perceived self-orientation with and without these auditory "spatial" cues may provide evidence that the pitch-height association could be used as a self-orientation cue.

Traditionally, our perception of "up" has been measured using the subjective visual vertical (SVV, Aubert, 1861; Howard, 1982) by having subjects judge or set the orientation of a visual line with the perceived direction of gravity (i.e., the direction a physical object would fall). A recent and complementary concept of "up" is the perceptual upright (PU, Dyde et al., 2006), defined as the orientation at which an object or character is most easily recognized. The PU can be measured by having subjects judge

whether they identify the character ‘p’, presented in various orientations, as either the letter ‘p’ or the letter ‘d’ in a task called the Oriented Character Recognition Task (OCHART). By finding the average angle of the transition points between ‘p-to-d’ and ‘d-to-p’ the PU can be determined from the point in between these “maximum ambiguity” orientations. Both representations of “up” are influenced by the different frames of reference that the brain uses to determine upright. These reference frames include the idiotropic vector (or body vector), the gravity vector (which is built up from vestibular and other cues, Lackner & DiZio, 2005), and the visual vector (based on visual cues to upright such as the light-from-above prior and the horizon line, Dyde et al., 2006).

These relative influences on a participant’s SVV and PU can be modeled using a vector sum model, first conceived of by Mittlestaedt (1983) and further developed by Dyde et al. (2006). In this model (Figure 6) the length of each vector represents the relative weighted influence of each frame of reference. The model for the SVV typically shows a very large influence of the gravitational vector and little else while the PU shows a more evenly distributed set of weighted influences. Thus, an advantage of modeling the PU over the SVV is that it is a more sensitive measure for determining the relative influence of the various contributors to the sense of upright. Thus, presenting participants with the OCHART and detecting any difference in the PU as a function of the presentation of auditory cues, could reveal changes to the underlying representation of self-orientation due to the auditory cue. Furthermore, any effect of sound on the PU could be modeled as an added vector specific to the influence of sound (Fig 5B).

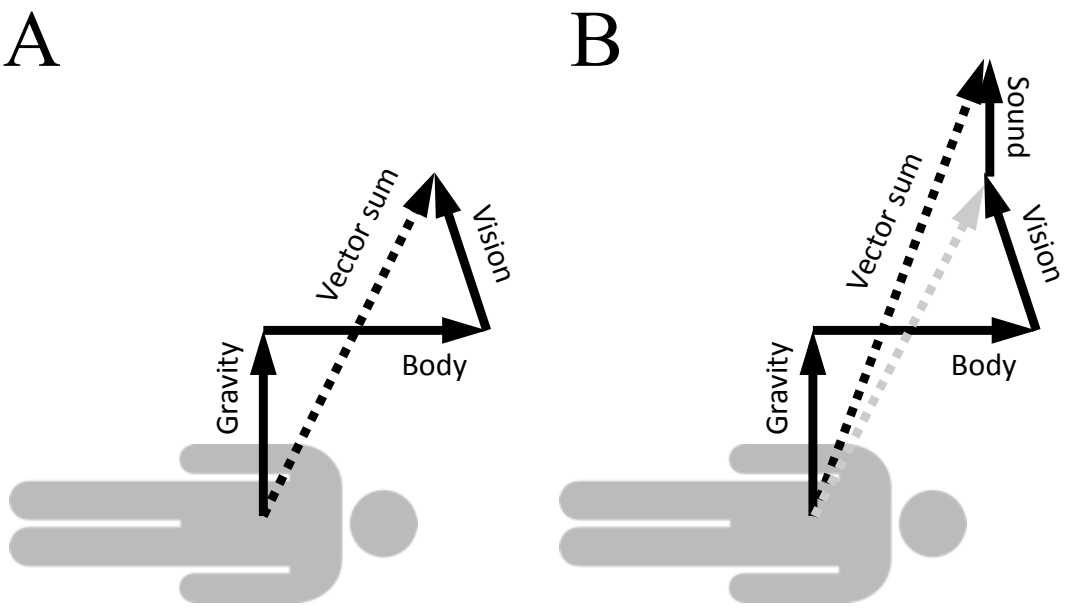

Figure 6. (A) Vector sum model for perceptual upright with human silhouette lying on their right side. The solid black arrows represent the relative influence of each reference frame and the black dotted arrow represents the perceptual upright. (B) Vector sum model for perceptual upright incorporating the potential influence of a sound cue signaling "upright" in the direction of gravitational up. The two dotted lines show the difference in the perceptual upright when the sound cue is present (black) and not present (grey).

There is very little research specifically looking at the role of sound as a cue to perceived self-orientation. Somewhat related to perceived self-orientation, Väljamäe (2009) presented a review article on the effects of auditory stimuli on vection. Vection refers to perceived illusory self-motion and is most commonly created by visual stimuli. The author concluded that while auditorily induced self-motion perception is weaker than

visually induced self-motion perception, specific acoustic cues could be useful for self-orientation domains such as posture prosthesis, navigation in unusual gravitoinertial environments, non-visual navigation, and multisensory integration during self-motion. While these forms of perception related to one's position in space are not perceived self-orientation per se, these measures are the closest analogues available in the literature. One such example is from Easton et al. (2008), where they demonstrated that by presenting blindfolded standing subjects with static noise from speakers positioned adjacent their left and right ears, postural and head sway were minimized. In this study, participants were able to use the static sounds as a continued reference for their bodies' posture. While various auditory properties (i.e., amplitude, sound source position, sound dynamics, head-related transfer functions etc.) and their applicability to perceived self-orientation have been researched to some extent there is no work that looks specifically at the potential of the crossmodal correspondence between pitch and visuospatial elevation.

To test whether the auditory pitch-height correspondence can play a role in self-orientation, participants performed the OCHART with and without sounds that were strategically placed to exploit the association. Participants laid on their right side and speakers were placed above and below the participants' left and right ears respectively. The auditory stimulus was composed of a high tone above and a low tone below participants' heads. The aim of introducing this stimulus was to present an auditory cue that would increase the perceptual weighting of the reference frame oriented along the gravitational axis. This should then pull the PU further towards this orientation (see Figure 6B). To be clear, the auditory cues presented in the experiments in chapter 2 include both a true spatial component (i.e., the speakers are unambiguously localized in

physical space by the participant as above and below their heads along the gravitational axis using binaural cues) and a crossmodal correspondence component (i.e., the high and low pitched tones presented from the above and below speakers, respectively), which together might serve as a cue to gravitational "up" and "down".

1.6.2 Experiment 2 – What reference frames are used in binding auditory pitch to spatial elevation for visual motion perception?

In experiment 2, the reference frames used in determining "up" for binding ascending and descending auditory tone-sweeps to visual motion were investigated. High and low tones have been shown to bind perceptually to high and low spatial elevations respectively, but what constitutes "high" and "low" visual space? As mentioned in the previous section (1.6.1), "up" can be defined in reference to the axis of the body, gravity, the visually defined scene, or a weighted combination of these reference frames. In Maeda et al. (2004), multisensory integration between ascending and descending tones and ambiguous visual motion gradually broke down as the axis of visual motion was tilted from upright. In their experiment, where participants sat upright, the gravity and body reference frames were aligned however, so whether "upright" was relative to the body or to gravity is unclear. Here I ask: which of these reference frames play a role in binding ascending and descending tones to "upward" and "downward" visual motion?

While Parise et al. (2014) investigated the reference frames used in the localization of sounds the present study is different in important ways. Unlike the Parise et al. (2004) study, which looked at how real, spatially laid out sounds were localized in blank visual space, this study aims to investigate the parameters by which an auditory

stimulus can perceptually bind to a visual stimulus.

To test which reference frames are used in binding ascending and descending tones to visual motion stimuli, participants performed a modified version of the visual motion discrimination task from Maeda et al. (2004) but while participant's body orientation and the orientation of visual motion were manipulated. Participants performed the discrimination task in four experimental conditions with two body orientation conditions (upright vs. lying on right side) and two visual motion direction conditions (upward/downward motion vs. leftward/rightward motion with respect to head). In this way, the gravity and body reference frames can be decoupled and the degree to which the sounds influenced visual perception can be compared. To illustrate, if sounds are found to bias visual motion perception while participants are on their right side and participants are making a leftward/rightward judgment (i.e., visual motion is along the gravity axis only), than this would suggest that the brain uses the direction of gravity as a reference to "up" for binding the audiovisual information.

In this experiment, the possibility of a response bias influencing the results is also tested. While there appears to be consensus in the literature that the pitch-height association can lead to true multisensory integration, the relative lack of studies pertaining to visual motion compared to other tasks leaves the interpretation of this effect unclear. Could the presence of a sound merely bias a participant's tendency to respond in a particular direction rather than actually lead to a bias in visual perception? Thus, a series of response bias control trials were interleaved into the experiment where the auditory and visual stimuli were temporally asynchronous. In this way, any effect of sound in these trials would suggest that participants are responding primarily to the

sound, reflecting response bias, and not to the visual stimulus.

1.6.3 Hypotheses

The main hypotheses are as follows:

Experiment 1 Hypothesis: Low and high tones presented strategically as auditory cues to gravitational upright will bias participants' perceptual upright, as measured using the oriented character recognition task, compared to a no-sound condition. This would suggest that the pitch-height crossmodal correspondence can be used as a self-orientation cue.

Experiment 2 Hypotheses: Ascending and descending tones played through headphones will bias the perception of ambiguous visual motion when the directional axis of visual motion is aligned with either the head or gravitational spatial reference frames. My hypothesis is that visual motion perception will be biased by sound in all orientations of visual motion direction conditions except when participants are upright and making leftward/rightward visual motion judgments (i.e., the only condition where visual motion is not travelling along either gravitational or bodycentric upward and downward). A secondary hypothesis is that the ascending and descending tones will not influence visual motion perception in trials where auditory and visual stimulus presentation is asynchronous (i.e., high and low tones will not, introduce a response bias).

2. CAN HIGH AND LOW AUDITORY TONES INFLUENCE THE PERCEPTUAL UPRIGHT?

2.1 Overview

Here, I investigated the potential for the pitch-spatial elevation crossmodal correspondence to play a role in spatial orientation. In two related experiments, participants laid on their right sides, thus setting the gravitational and bodycentric spatial reference frames in conflict, and were presented with sounds intended to act as a spatial reference cue to gravitational upright. Participants performed the OCHART using an adaptive staircase paradigm (see section 1.6.1 for theoretical description of the OCHART and PU), to measure perceived self-orientation in a control condition with no sounds present and in an experimental condition with sounds present. My hypothesis is that when the sounds are present, participants' OCHART results will reflect a greater influence of the gravitational vector on their perceptual upright.

In the first experiment, static sounds were used and in the second experiment a more complex dynamic sound stimulus was used. In the static-sounds experiment high and low tones were simultaneously presented via speakers above and below the participants' heads (in gravitational terms) respectively. In the dynamic-sounds experiment, a more complex auditory stimulus was used where a low-frequency sound began below the participant's head at low volume in the bottom speaker and increased in volume and frequency as it transitioned to the above speaker. This gave the impression of a sound travelling from below the participant's head at low frequency, "through" the participant's head, to above the participant's head with a high frequency. The concept of

a “sound image” being perceived as moving through the participant’s head was inspired by work from Lewald and Ehrnstein (1998) who used sounds of relative volume to measure the perceived spatial location of a sound image under various body orientations. Furthermore, Zohar Eitan, Schupak & Marks (2010) demonstrated a crossmodal correspondence between visual height and auditory volume such that higher volume sounds were associated with higher visual space. The dynamic-sounds condition was therefore an attempt to combine these effects to enhance the perceptual influence of the auditory spatial cue.

2.2 Methods – Static tones

2.2.2 Participants

Twelve participants (6 male, 6 female, mean age 25) volunteered to participate in this experiment. These participants were all graduate students. All participants reported having normal or corrected-to-normal vision and no known vestibular or self-orientation issues. All experiments were approved by the ethics board of York University and followed the guidelines of the Declaration of Helsinki.

2.2.3 Apparatus and general setup

Participants performed this experiment laying on their right side on a massage table in a large dark storage room. The room’s dimensions were approximately 5 by 5 metre length walls 2 stories high with a concrete floor and wood ceiling. The participant’s head lay off the edge of the massage table and was held parallel to their body axis by a custom support apparatus that ensured that their right ear was unoccluded

and exposed towards the ground and left ear unoccluded and exposed to the ceiling A computer monitor (ViewSonic VG732M-LED, display size: 13” horizontal x 10.6” vertical, resolution: 1280x1024, pixels-per-inch: 96.2) was aligned with the participants’ head. Participants viewed the monitor through a circular black shroud (diameter 35°), which kept their head at a constant distance of 20 cm from the display and blocked out external visual stimuli. Speakers were positioned above and below the participant’s head with the top speaker aimed down to the participant’s left ear and the bottom speaker aimed up at the participant’s right ear. Both speakers were 54 cm from the participant’s head. Participants held a computer mouse in their right hand with which they made responses. The monitor, speakers, and mouse were all connected to a MacBook Pro and all inputs and outputs were controlled using MATLAB. All visual stimuli for all experiments were generated and presented using the Psychtoolbox package of functions (Brainard, 1997). Anti-aliasing features were utilized for visual smoothing of all visual objects. All auditory stimuli were presented using the Psychportaudio MATLAB package functions.

2.2.4 Visual stimulus and adaptive staircase

In implementing the OCHART, participants were presented with the ambiguous visual character “p”, which can be interpreted as a “p” or a “d” depending on its perceived orientation. The character, presented as a capital “P” in Calibri font subtended 5.75° X 4°. To obtain the participant’s perceptual upright, the two orientations where the character is most ambiguous must be determined. To determine these two orientations, two QUEST adaptive staircases were used (Watson & Pelli, 1983) to adjust the orientation of the presented characters to find the threshold orientations where

participants reported a “p” or “d” 50% of the time. The PU is calculated by finding the point midway between the two orientations at which the p/d character is most ambiguous. The QUEST staircase uses a Bayesian method to estimate the threshold value. The two staircases began with the character at opposing orientations of -90° and +90° (positive angles representing rotation in the clockwise direction and counterclockwise rotation for negative angles) where 0° was defined as the top of the participant’s head (Figure 7A). The two staircases were programmed such that when participants responded that they saw a “p”, the presented character that was initially tilted over at -90 degrees orientation would likely be tilted further counter-clockwise on the next presentation while the character initially presented at 90 degrees would likely be tilted in the clockwise direction and vice versa (Figure 7B). All staircases terminated after a pre-set number of presentations (Figure 7C, also see section 2.2.6 and 2.3.3 for number of trials).

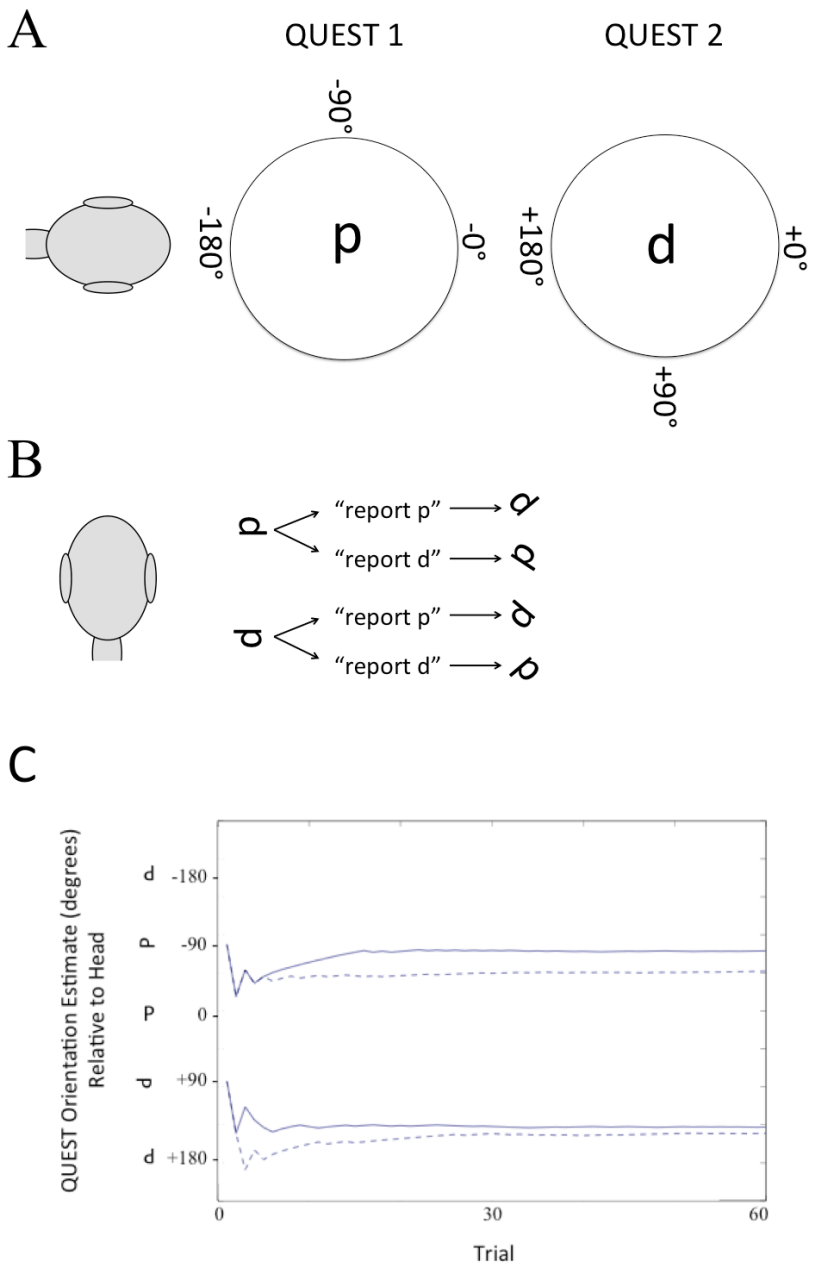

Figure 7. OCHART presentation and QUEST adaptive staircases. (A) The tilted participant and the presentation of the "p/d' stimulus. For each auditory condition there were two staircases, one starting with an orientation of +90 degrees and the other with -90 degrees. (B) Simplified demonstration of QUEST staircase orientation adjustments based on participant responses (head centered coordinates). Note the algorithm re-orients

the probe in such a way that a report of "p" leads to a character that leads to a "d", and vice-versa. Through successive iterations, this principle leads to the point where the probe is most ambiguous (i.e., equally likely to be identified as a p or a d). (C) Example staircase data taken from a typical participant's performance which shows probe orientation by trial. The dotted lines represent the participant's responses with sounds present while the solid lines represent trials with sounds absent. As this participant response data shows, the staircase appears to converge on the PSE fairly quickly.

2.2.5 Auditory stimulus – Static sounds

The auditory stimulus for the static-sounds experimental condition was composed of pure tones where the speaker located above the participant's head played a pure tone of 1200 Hz and the speaker below simultaneously played a pure tone of 200 Hz, lasting for 400 milliseconds (Figure 8). During control trials, no auditory tone was presented. For all experiments described in this thesis, a sampling frequency of 44100 Hz was used to generate and play all sounds auditory stimuli.

2.2.6 Experimental paradigm

The within-subjects experimental paradigm was composed of two auditory conditions: the experimental condition (pure tones) and control condition (no sound). To obtain the PU each participant completed two adaptive staircases for each sound condition. Each staircase was set to run for 60 trials leading to a total of 240 trials (2

conditions X 2 adaptive staircases X 60 trials each). The presentations of all four staircase conditions were randomly interleaved in the experimental paradigm.

The experimental procedure was as follows. Every trial began with the "p/d" probe which was presented for 400 milliseconds in a particular orientation chosen by the adaptive staircase (except the starting orientations which were pre-set, see Figure 7) depending on the participant's previous responses. During experimental trials the auditory stimulus was presented with the visual stimulus for the duration of the probe stimulus while in control trials no sounds were presented. After the probe disappeared participants responded after an enforced 300-millisecond delay. A left mouse-click denoted that they perceived a "d" and right mouse-click denoted a "p". After their response there was a 400 ms delay before the next trial was presented. This is shown in Figure 7B.

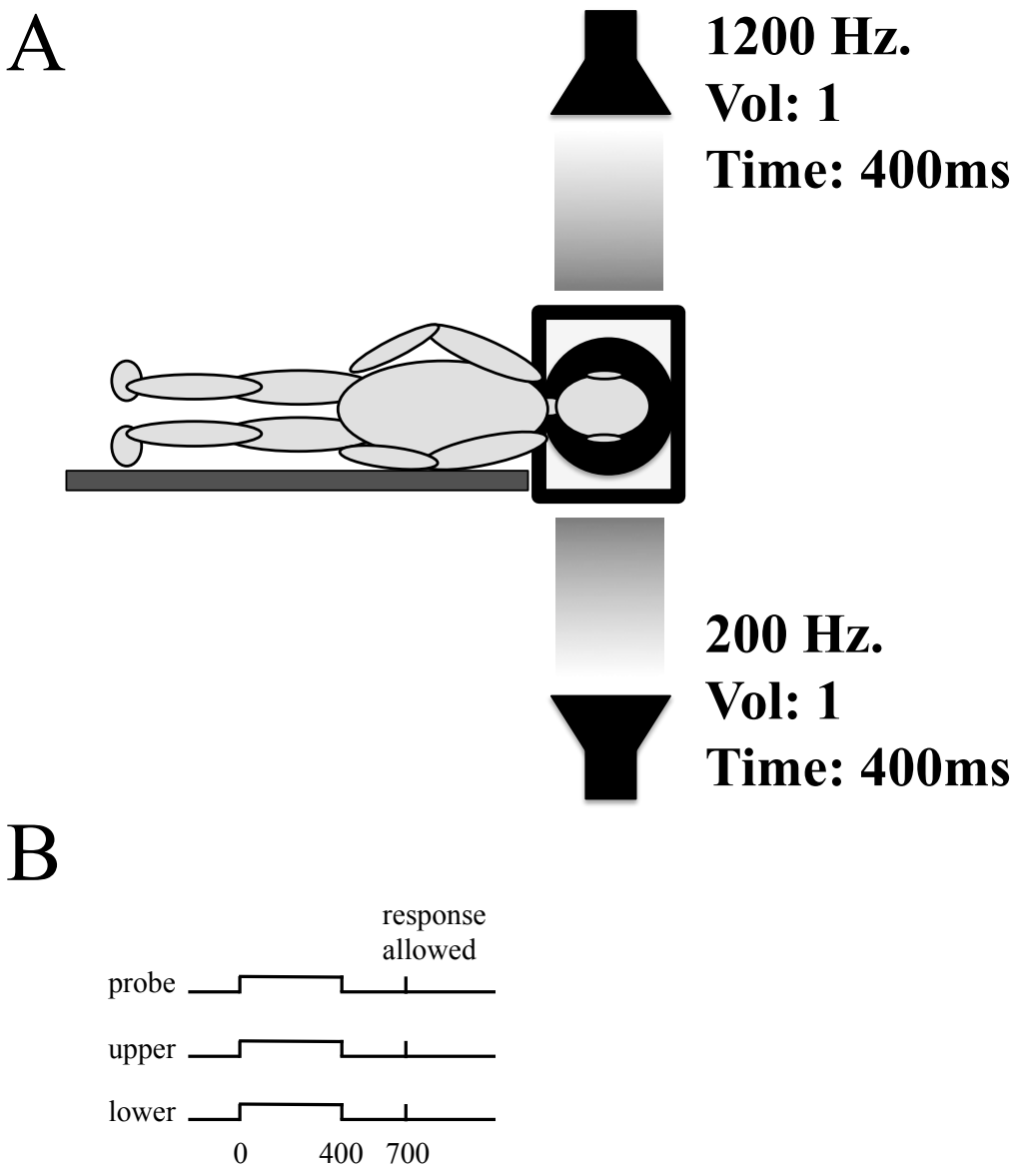

Figure 8. Auditory stimulus for the static-sounds experiment, overall setup, and trial timings. (A) Participants laid on their right side on a support surface with their ears unoccluded. Participants' viewed the screen through a circular shroud. Speakers above and below the participant's head played sounds simultaneously as shown. (B) Trial timings for the "p/d" probe stimulus, auditory stimuli from the upper and lower speakers, and delay period before responses can be made in milliseconds.

2.3 Methods – Dynamic tones

2.3.1 Participants

Thirteen participants (5 male, 8 female, mean age 24) volunteered to participate in this experiment. These participants were graduate students and students who volunteered from the Undergraduate Registered Participant Pool (URPP) who were awarded class credit for participation. All participants reported having normal or corrected-to-normal vision and no known self-orientation or auditory issues. All experiments were approved by the ethics board of York University and followed the guidelines of the Declaration of Helsinki.

2.3.2 Auditory stimulus – Dynamic sounds

The auditory stimulus from the experimental condition was composed of a tone that began with the bottom speaker and ended with the above speaker and changed its frequency and volume as it did so. The sound from the bottom speaker swept from 0 Hz to 600 Hz over a period of 500 milliseconds and immediately following this the above speaker played a tone sweep from 600 Hz to 1200 Hz over a period of 500 milliseconds (Figure 9). This lead to the perception of an auditory "object" rising both physically (from low space to high space travelling through the head) and in frequency (from 0 Hz to 1200 Hz) over a period of 1 second. The whole auditory sweep vector as was calculated in MATLAB (i.e., the two sweeps combined as one array) was multiplied by a ramp from 0-1 leading to a linear increase in volume as the sound "travelled" from low to high physical space. This was done to make the sound appear more as an object

perceptually and add to the sense that it is travelling from low to high space (see section 2.1 for more on the dynamic-sounds stimulus).

2.3.3 Experimental paradigm

Like the static-sounds experiment this was a within-subjects experimental paradigm composed of two auditory conditions: the experimental condition (dynamic-sound stimulus) and control condition (no sound). To obtain the PU under both sound conditions each participant completed 2 adaptive staircases each. Each staircase was set to run for 50 trials leading to a total of 200 trials (2 conditions X 2 adaptive staircases X 50 trials each). The number of trials per staircase was reduced in this experiment because it was found in the static-sounds experiment that the estimates converged within this number of trials (Figure 7C). The presentations of all four staircase conditions were randomly interleaved in the experimental paradigm.

The experiment was carried out the same as the static-sounds experiment but with a different auditory stimulus and changes to the presentation timings. Each trial began with the 1-second auditory stimulus. During playback at the 600-millisecond mark the visual probe was presented. After an additional 400 milliseconds both the visual probe and auditory stimulus ended and the screen went grey. During control trials there was a 600-millisecond delay after which the character probe was presented for 400 milliseconds followed by grey screen. After the greyscreen was a 500-millisecond delay after which participants' clicks registered their response, and the next trial would begin. This is shown in Fig 8B.

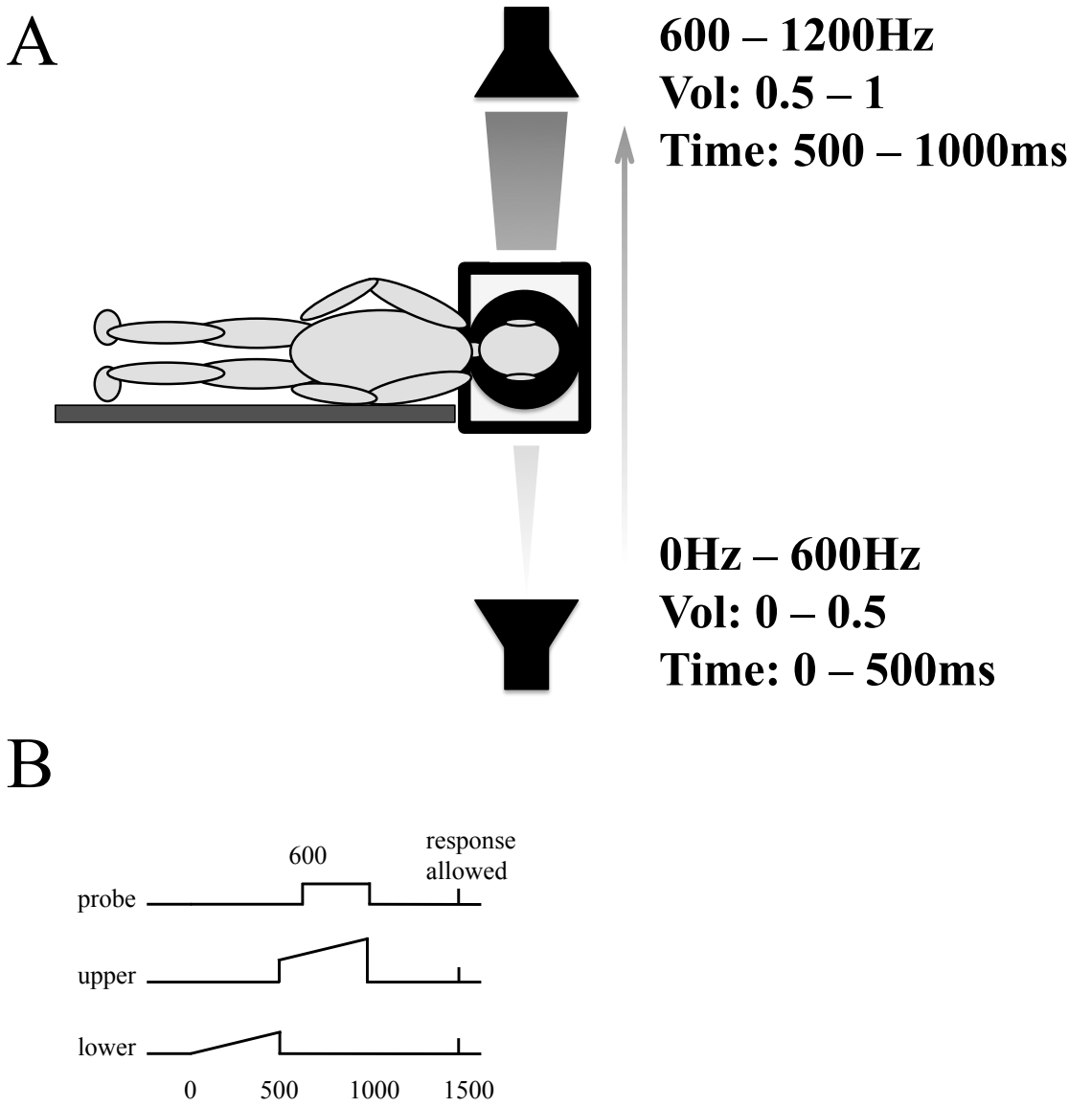

Figure 9. Auditory stimulus for the dynamic-sounds condition and trial timings. (A) The sound stimulus began with the bottom speaker at 0 Hz and increased in frequency to 600 Hz over a period of 500 ms. The auditory stimulus then continued in the above speaker which increased from 600 Hz to 1200 Hz for 500 ms, meaning the entire auditory stimulus lasted for 1 second. The entire auditory stimulus increased in volume linearly from 0 to 1. (B) Trial timings for the auditory stimulus, probe stimulus, and onset of the response period in milliseconds.

2.3.4 Data Analysis

For both the static- and dynamic-sound experiments the results were determined in the same way as follows. To determine if the PUs of the sound and no-sound conditions were different, first the points of subjective equality (PSE) had to be determined. The PSEs represent the orientations of the p/d values that were most ambiguous to participants. The PSE values from the four adaptive staircases (QUESTs starting at both -90 and +90 degrees orientation for each of the sound and no-sound conditions) were obtained by using the "QUESTMean" function from the MATLAB package developed by Watson and Pelli (1983). The QUESTMean function provides the final estimate of the routine calculated as the mean of the posterior distribution function and represents the best estimate of PSE. Second, to compare the values of the PUs from the sound- and no-sound conditions they had to first be calculated using the above PSE values. The PU is defined as the orientation midway between the two corresponding PSE values, thus calculating the mean gives the PU. For each of the PSE and PU measures as determined, standard errors were then calculated across subjects for each. To determine if there were any effects of sound on participants' PUs, within-subjects t-tests were performed comparing the sound and no-sound conditions.

2.4 Results

Static-Sounds. PSE and PU data are shown in Figure 10. For the no-sound condition the staircase starting at +90º resulted in PSE values with mean=67.4º and standard error=5.7º. The -90º mean=-119.1º with standard error=6.5º. For the sound

condition the staircase starting at +90º resulted in PSE values with mean=76.7º and standard error=7.0º. The -90º mean=-118.8º and standard error=7.0º.

The PU for the no-sound condition was -25.8º with standard error=4.8º and the PU for the sound condition was -21.1º with standard error=6.3º. The absolute difference between the PUs from the sound and no-sound condition is 4.8º towards the top of the head (when sounds were presented). A within-subjects t-test revealed no significant difference in PU values between the sound and no-sound conditions with t(11)=-1.57, p=0.14, d=-0.24.

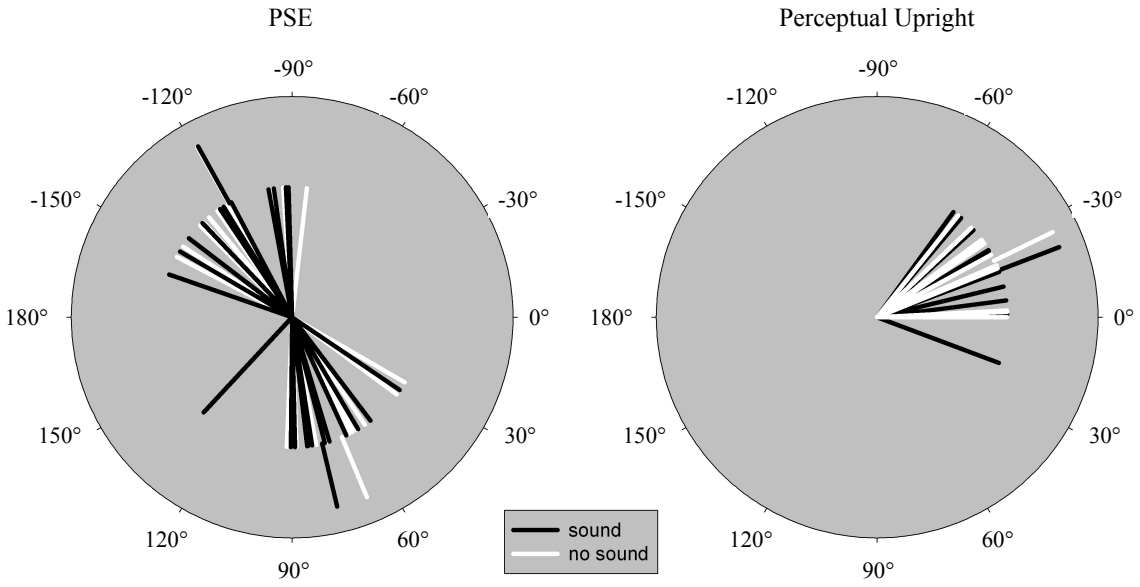

Figure 10. Polar plots showing PSE and PU values for the sound (black) vs. no-sound (white) conditions in the *static-sounds experiment*. 0º corresponds to the top of the participant's head and -90º corresponds to the gravitational up. Left: Inner lines show the raw data for the PSEs while the outer lines show mean values for the sound and no-sound conditions (note: the means for the sounds and no-sound conditions fall on top of each other at approximately -120º). Right: Inner lines show the PUs of each subject (calculated

from the PSEs shown in the left plot) and the outer lines represent the means of the PUs from the sound and no-sound conditions.

Dynamic-Sounds. PSE and PU data are shown in Figure 11. For the no-sound condition the staircase starting at +90º resulted in PSE values with mean=66.1º with standard error=5.5º. The -90º mean=-126.9º with standard error=8.5º. For the sound condition the staircase starting at +90º resulted in PSE values with mean=65.2º with standard error=5.2º. The -90º mean=-128.9º with standard error=8.71º.

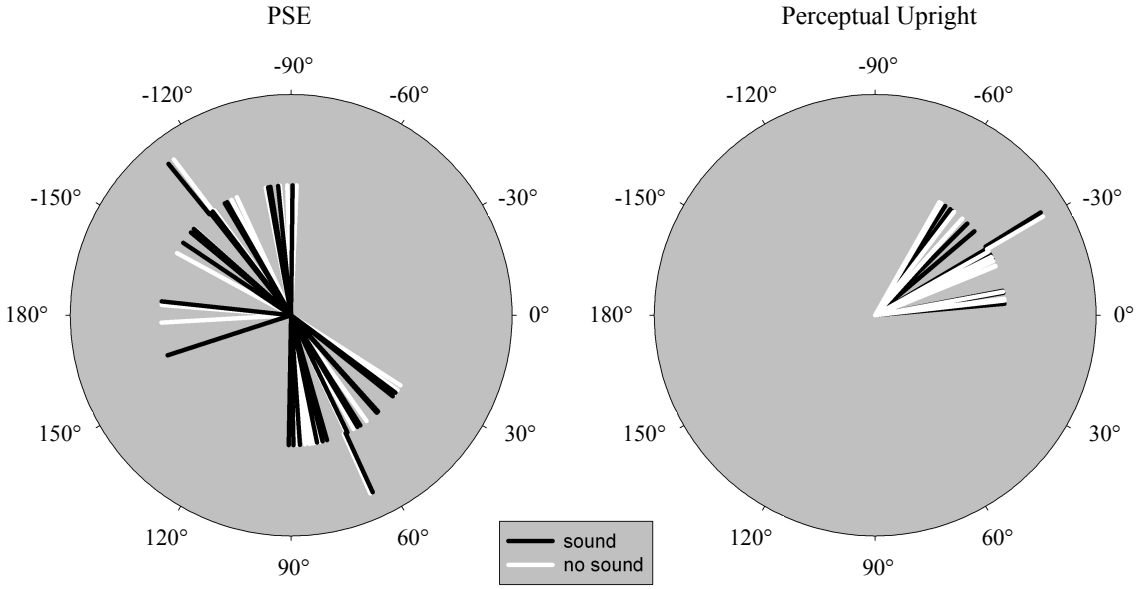

Figure 11. Polar plots showing PSE and PU values for the sound vs. no-sound conditions in the *dynamic-sounds experiment*. Format as for figure 10.

The PU for the no-sound condition was -30.4º with standard error=5.2º and the PU for the sound condition was -31.9º with standard error=4.8º. The absolute difference in PUs from the sound- and no-sound conditions is 1.4º. A within-subjects t-test revealed no significant difference between the PUs of the sound vs. no-sound conditions with $t(12)=-0.72$, $p=0.48$, $d=0.08$. In summary, in neither the static nor the dynamic-sound conditions did the presentation of sounds have a significant influence on participants' perceptual upright.

2.5 Discussion

In both the static- and dynamic-sounds experiments, within-subjects t-tests revealed that there were no significant differences between the sound and no-sound conditions. There was no tendency to strengthen the gravity vector (Fig 5) and swing the PU towards the gravitational vertical (towards the -90º). This suggests that the crossmodal correspondence between frequency and spatial elevation does not contribute to perceived self-orientation. This may not be entirely surprising as auditory cues in vection (Väljamäe, 2009) and posture (Easton et al., 1998) studies show that while small influences of sound have been demonstrated, visual stimuli are much more effective cues for vection and balance. Thus, even if sounds can potentially play a role in perceived self-orientation, their effects were too small to be detected in our experiments. It may be worth noting that for the static-sounds experiment the p-value was 0.14, which may possibly reflect a trending influence of sound. This is compared to the indubitably insignificant dynamic-sounds experiment where $p=0.48$. Similarly, the effect size (i.e., absolute difference in PUs between sound and no-sound conditions) in the static-sounds experiment is 4.8º (towards the head when sounds were presented) compared to 1.4º

(towards gravity) in the dynamic-sounds experiment, which is 3.4º larger. Standardizing the effect sizes (i.e. Cohen's d) similarly shows a larger effect in the static-sounds experiment (d = -0.25) compared to the dynamic-sounds experiment (d = 0.08).

The results from these experiments are consistent with Dyde et al.'s (2006) OCHART results. The PUs obtained from the static- and dynamic-sounds experiments are in agreement to where participants laid on their sides. The mean of the four PUs collected across the two experiments was -27.3º and the mean PU from Dyde et al. (2006) where participants laid on their right side was -17º. The PU was thus in between "up" defined by the body (0º) and the "up" defined by gravity (-90º or 270º), which is within the range reported by Dyde et al. (2006). The present findings thus demonstrate that using the QUEST adaptive staircase method to measure OCHART is psychophysically sound and leads to similar results found previously when using the method of constant stimuli (Dyde et al., 2006).

There are certain issues in experiment 1 that should be addressed. One concern of this study is that perhaps the number of participants was not enough to detect an actual hidden effect, due to insufficient power. As shown above, the static sounds experiment had a p-value of 0.14 and comparing their standardized effect sizes shows a noticeably larger effect in the static-sounds (d = -0.25) and dynamic-sounds (d = 0.08) experiments. The second issue is that in designing the auditory stimulus, certain acoustic features such as reverberation (i.e., sounds from the speakers could have interacted with the surrounding environment to confound the cue, Waterhouse, 1958) or volume-frequency interactions (i.e., the sounds may not have been perceived as equally loud, Robinson & Dadson, 1956), were not specifically taken into account, leading to possible confounds

and/or less than optimally effective stimuli. However, these are likely to be very small effects because of the proximity of the speakers to the participant's head, which should have been the most prominent auditory cue. Another potential shortcoming of this study is the fact that only a select range of frequencies was used, despite the fact that Parise et al. (2014) showed that different frequencies are associated with different spatial elevations (see Figure 3). Perhaps sounds of other frequencies not tested would be used as auditory cues relevant to perceived orientation.

In summary, the results suggest that the crossmodal correspondence between auditory pitch and spatial elevation does not play a role in determining the direction of perceived self-orientation. The results confirm previous OCHART studies and these experiments show that the QUEST adaptive staircase technique can be used reliably to measure the PU. Put another way, this also shows that the PU as a measure is robust to different methods of measurement. Finally, perhaps flaws in the experiments' design washed out possible effects, and remedying these issues might reveal that this crossmodal correspondence can in fact be used as a spatial cue to perceived self-orientation.

3. WHAT REFERENCE FRAMES ARE USED IN INTEGRATING ASCENDING AND DESCENDING TONES WITH AMBIGUOUS VISUAL MOTION?

3.1 Overview

Here, I investigated the spatial reference frames used in combining auditory pitch and visual motion stimuli. This experiment was done to reveal whether the brain uses the gravitational, bodycentric, or a combination of both reference frames when perceptually binding "upward" and "downward" visual motion with ascending and descending auditory tones. This experiment is an extension of work by Maeda et al. (2004), which showed that upward/downward ambiguous visual motion discrimination was biased by the presentation of ascending and descending auditory tones. In Maeda et al. (2004) however, participants performed this task while sitting upright, so it is not known whether the effect depended on the gravitational, bodycentric reference frame, or a combination of the two.

In the present experiment, participants either sat upright or laid on their right sides in order to set the gravitational and bodycentric reference frames in conflict with each other spatially. In each position, participants were presented with a more or less ambiguous visual motion stimulus nearly identical to that of Maeda et al. (2004) but in this experiment the axis of visual motion was either upward-and-downward with respect to the participant's head, or leftward-and-rightward. This created four combinations of body orientation and visual motion directions. The combinations were: upright upward/downward visual motion, upright leftward/rightward, on-side upward/downward, and on-side leftward/rightward.

Participants were asked to discriminate whether they perceived predominantly upward or downward visual motion or leftward or rightward visual motion depending on the experimental condition. In the on-side leftward/rightward condition, visual motion is along the gravity axis only, and therefore if ascending and descending tones influence the perceived direction of visual motion, it demonstrates that the stimuli are perceptually bound along the gravitational axis. In the on-side upward/downward condition, visual motion is only along the body axis, and therefore effects of sound would show that they are bound along the body reference frame. If there are effects of sound along both reference frames, it would suggest that both play a role. By determining in which conditions there is a biasing effect of sound on visual motion perception (found in Maeda et al.'s study), the reference frame used in binding auditory frequency and visual motion stimuli can be revealed.

When identifying the direction in which an inherently ambiguous visual stimulus appears to move, there is a potential confound of response bias. That is, the presence of a sound cue may itself encourage the observer to make one or other choice rather than by true perceptual bias. Special trials were therefore interleaved within each of these experimental conditions to test for a potential response bias. By presenting the tones *asynchronously* from the visual motion stimulus it can be inferred whether or not participants are responding to perceived visual motion (which should no longer bind with the asynchronous sound) or if their response was biased by the presentation of the sound stimulus. The auditory stimulus was presented afterwards so participants could perceive the visual stimulus in isolation, where any biasing effect of sound would be due to an influence at the decision level.

3.2 Methods

3.2.1 Participants

Nine participants (5 male, 4 female, mean age 25) volunteered to participate in the four within-subjects experimental conditions. All participants reported having normal or corrected-to-normal vision and no known vestibular or self-orientation issues. All experiments were approved by the ethics board of York University and followed the guidelines of the Declaration of Helsinki.

Before participating in the study, participants performed a short practice regime to ensure that they could in fact perform the task and thus provide usable data. Nine participants could not properly perform the task and thus did not participate in the experiment. See the end of section 3.2.5 for more details on the procedure of the practice regime.

3.2.2 Apparatus and general setup

Participants performed the experiment both upright and laying on their right side in a small dark room. In the upright conditions, participants sat comfortably on a chair at a table and looked at a monitor (ViewSonic VG732M-LED, display size: 13” horizontal x 10.6” vertical, resolution: 1280x1024, pixels-per-inch: 96.2) through a black circular shroud (diameter 35°), which kept their head at a constant distance of 20 cm from the display and blocked out external visual stimuli.

During the on-side conditions participants laid on their right side on a padded massage table and looked at an identical monitor and shroud, which were tilted with the

participant. Padding was offered to subjects to make them more comfortable. Auditory stimuli were presented via noise-cancelling headphones (Maxell NC-11) in both the upright and on-side orientations. Participants held a computer mouse in their right hand, which was used to make responses. The monitor, speakers, and mouse were all connected to a MacBook Pro and all inputs and outputs were controlled using MATLAB.

3.2.3 Visual Stimuli

The visual stimuli were composed of two superimposed, spatially enveloped sinusoidal luminance gratings with a Michelson contrast of 0.05, spatial frequency 2 cycles/degree, and temporal frequency 6.25 Hz drifting in opposite directions (grating speed is thus 12.5 degrees/second). The gratings filled the 35º circular aperture. To create ambiguous motion, the two component gratings were presented with various contrast ratios (upward/downward or leftward/rightward components: 1.0/0.0, 0.7/0.3, 0.6/0.4, 0.5/0.5, 0.4/0.6, 0.3/0.7, 0.0/1.0, see Figure 12). The gratings with 0.5/0.5 contrast ratio produced completely ambiguous motion (or flicker). Two visual motion conditions were run in separate blocks: horizontal and vertical movement relative to the observer.

3.2.4 Auditory Stimuli

There were three distinct sounds presented to participants in this study. Two of the auditory stimuli were tones with a constant rate of change, either ascending from 0.3 to 2.0 kHz, or descending from 2.0 to 0.3 kHz over a period of 200 milliseconds. The other auditory stimulus was broadband 1/f noise (i.e., noise signal with frequency spectrum such that the energy per Hz is inversely proportional to the frequency of the signal. Also known as “pink noise”; Bak, et al., 1987) presented for a period of 1 second.

Sounds were delivered through headphones. Presentation was the same in each ear and volume was constant throughout the experiment. Volume was selected for each participant to be loud enough but to not cause discomfort.

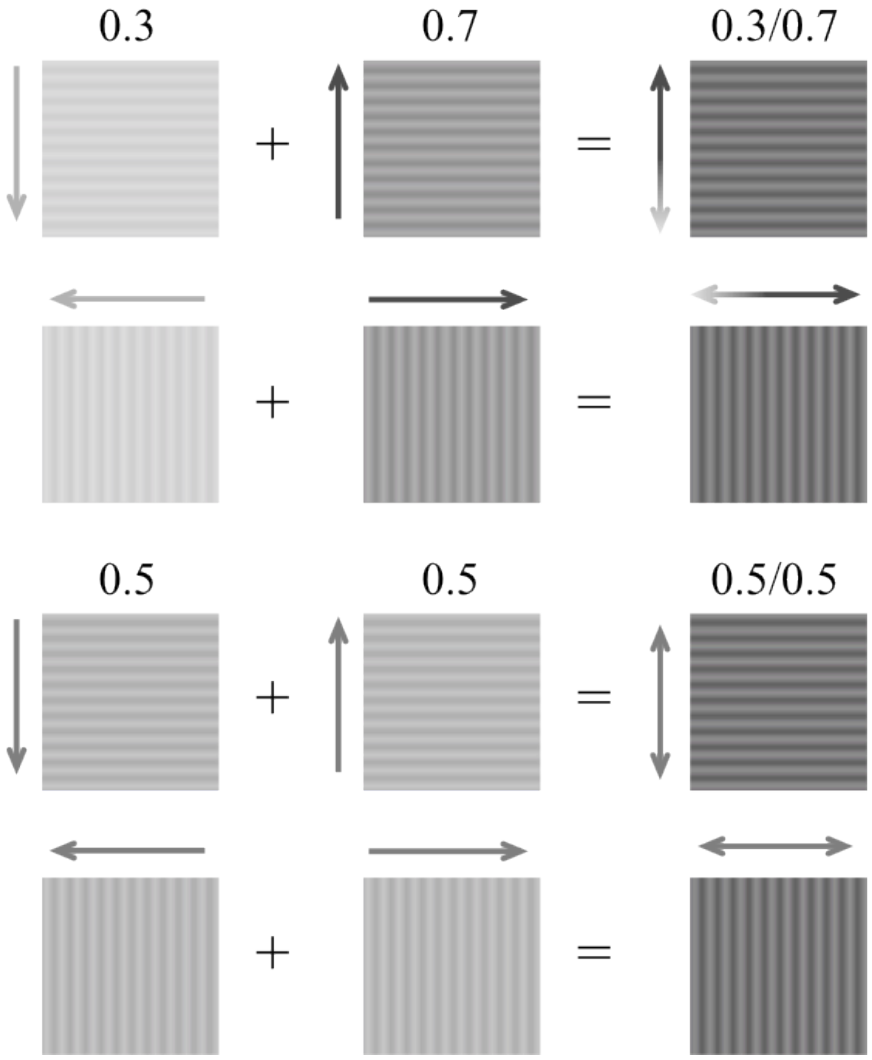

Figure 12. Diagram of the ambiguous visual motion stimuli used in this experiment. The visual stimulus was composed of two superimposed identical grating patterns moving in opposite direction at the same speed. The overlapped gratings had a variable contrast ratio and the resulting pattern thus appeared to move predominantly in one direction or the other, or completely ambiguously (when the contrast ratios were the same as shown

in the lower two rows). The top two rows show the superimposed gratings with contrast ratios 0.3 and 0.7 where the top row is the upward/downward stimulus and the second is the leftward/rightward. The bottom two rows show the ambiguous visual stimulus with contrast ratios at 0.5. Stimuli adapted from Maeda et al. (2004)

3.2.5 Experimental Paradigm

The within-subjects experimental paradigm was composed of four experimental conditions with two physical orientations (upright and lying down) and two directions of visual motion (horizontal and vertical relative to the observer), resulting in four within-subjects experimental conditions. Explicitly, these four conditions were: subjects sitting upright making upward/downward visual motion judgments, subjects laying on their right side making upward/downward visual motion judgments relative to their head, upright making leftward/rightward judgments, and on-side making leftward/rightward judgments (Figure 13A). All nine participants participated in all conditions in counterbalanced Latin-squares order.

Each trial began with a centered fixation cross of 2º on a grey background replaced after a randomized time delay of between 1-2 seconds with the moving gratings (duration 400ms) at one of the 7 contrast ratios chosen randomly. One of the three sounds (i.e., ascending, descending, or pink noise, duration 200ms) was played starting 50 ms after onset of the visual gratings stimulus. The visual motion stimulus was then followed by a grey-screen, which signaled participants to make a forced choice of the direction of motion they saw (i.e., either upwards or downwards, or leftwards or rightwards,

depending on the direction of visual motion condition). Relative to their head, if they saw downward or leftward visual motion they responded with a left-click of the mouse, and a right-click for upward or rightward motion. These trials are referred to as synchronous trials because the visual and auditory stimuli temporally overlapped (Figure 13B).

Mixed into the experimental design were a series of trials meant to detect possible response bias. In these trials only the three most ambiguous contrast ratios for the gratings were presented (0.4/0.6, 0.5/0.5, 0.6/0.4). One of two sounds (ascending or descending) was played 100 ms after the visual gratings stimulus presentation ended, and subjects responded after the sound ended. These trials are referred to as asynchronous trials because the auditory and visual stimuli did not temporally overlap (Figure 13B).

In each of the 4 orientation and visual motion direction combinations there were 20 synchronous trials for each of the 3 sound conditions at each of the 7 contrast ratios, leading to 420 trials. There were also interleaved 20 asynchronous trials for each of 2 sounds at 3 contrast ratios leading to a further 120 trials for each of the 4 combinations. The synchronous trials and the asynchronous trials were randomly interleaved, leading to a total of 540 trials in each of the conditions.

Before participants began the experimental procedure, they performed a practice regime to ensure that they could in fact do the task properly and thus give useful data. While sitting upright, participants performed condensed versions of the upward/downward and leftward/rightward motion discrimination tasks. They performed the task described above but in a condensed version where each trial condition was only performed once before the practice regime terminated (this included both synchronous

and asynchronous trials). All participants only performed the practice regime in the upright position. Participants were included to perform the actual experiment if they could demonstrate that they could properly identify the visual motion in the conditions with the least ambiguous contrast ratios (i.e., where the gratings were clearly moving upward/downward or leftward/rightward).

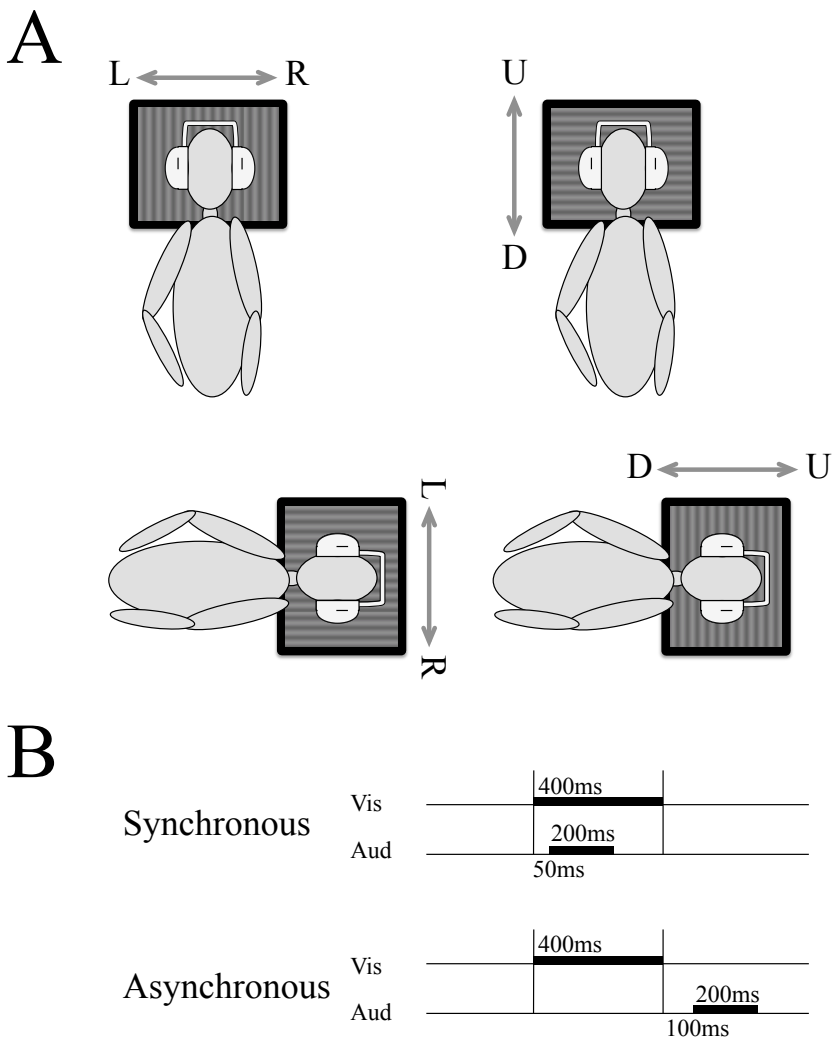

Figure 13. Experimental conditions and layout of the motion discrimination experiment. (A) The four orientation and visual motion combinations. In the top row, body orientation is upright whereas the bottom row demonstrates the on-side conditions. In the two left

columns the visual motion direction is leftward/rightward relative to head, and in the right two columns the visual motion direction is upward/downward relative to the head. (B) The timings of stimulus presentations for the synchronous and asynchronous trials.

3.2.6 Data Analysis

The effect of synchronous sound was evaluated by fitting a logistic function through each participant's data in order to obtain the points of subjective equality (PSE). A logistic function was chosen due to its symmetry since the contrast ratios of the visual stimuli were symmetrically graded from one direction of visual motion to the other. The function had 3 free parameters: α(PSE), β(slope), and γ(lapse rate) to fit. γ represents the difference value of both the top and bottom asymptotes from the minimum and maximum response values (0 and 20, then scaled to represent percentage). I chose to model a single lapse rate parameter to make the functions most comparable to each other while investigating how well the participants could perform the task at the extreme contrast ratios. The algorithm searched for the parameter values using a maximum likelihood optimization routine (Myung, 2003), which is the preferred method when dealing with generalized non-linear models such as this one (Fesselier & Knoblauch, 2006). The PSEs for this experiment represent the contrast ratio where participants were equally likely to report visual motion in either direction.

To explore the effects of the three independent variables (head orientation, direction of visual motion, and presence of sound) on participants' responses, the PSEs were inputted into a three-way analysis of variance. Before inputting the PSE data the values from the noise condition were subtracted from the values of the sound conditions.

These values were then inputted into a 2x2x2 within-subjects hierarchical mixed-models analysis of variance (ANOVA) tested using 3 different underlying co-variance structures to reveal which yields the best fit for the model (Field, et al., 2012). The covariance structure that yielded the best fit was the autoregressive covariance structure, which had the best measure of fit with the lowest value for the Akaike information criterion (AIC, Akaike, 1973) of -59.6, compared to unstructured (AIC=-58.1), and compound symmetric (AIC=-56.9) covariance structures. An advantage of using a mixed-models analysis over a traditional ANOVA is that some statistical power is gained while sphericity is controlled.

3.3 Results

Omnibus test

Figure 14A shows illustrative psychometric functions fitted through the mean data from all the participants. A 2x2x2 ANOVA (see 3.2.6) was performed on the PSEs obtained from each subject individually to explore the PSE data for the factors of body orientation (on-side or upright), direction of visual motion (upward/downward or leftward/rightward), and sound (ascending and descending, where the values from the. noise condition are subtracted from each). The ANOVA revealed a significant interaction between the effect of sound and visual direction with $F(1,56)=26.44$, $p<0.001$, generalized-eta-squared=0.36. The interaction between sound and visual direction revealed that averaging across orientation conditions, the effect of sound in the up-down tasks is greater than the effect of sound in the left-right tasks. There was no 3 way interaction with $F(1, 56)=0.38$, $p=0.54$, generalized-eta-squared=0.01.

The omnibus test however may not be the most appropriate way to explore these data and the hypotheses directly. The main reason for this is that the dependent variable across these conditions is not constant. Depending on the direction of visual motion presented, participants either made a leftward/rightward judgment or an upward/downward judgment. Thus it was not possible to properly put all four orientation by visual direction conditions into one consistent reference frame for comparison. To illustrate, in the on-side leftward/rightward condition, gravitational up corresponds to leftward responses whereas in the upright upward/downward condition, gravitational up corresponds to upward responses.

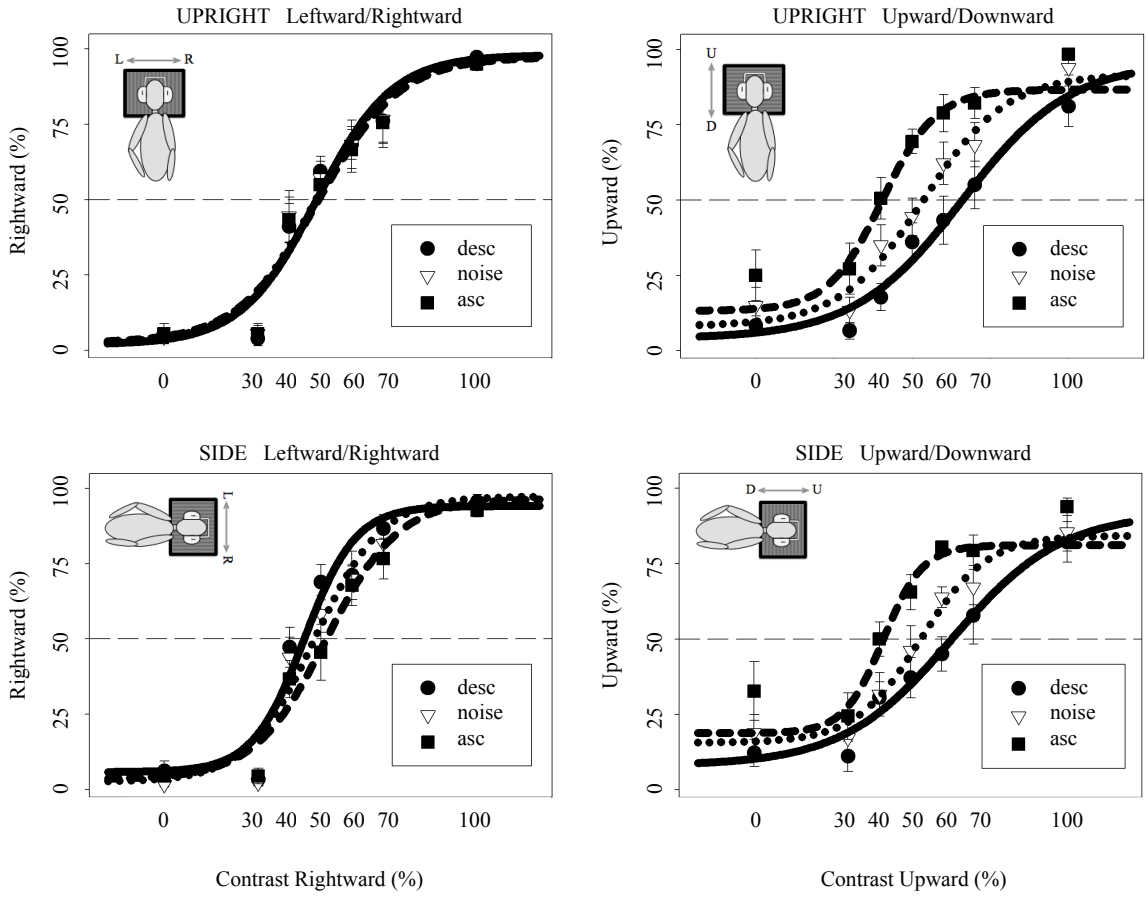

Figure 14. (A) Psychometric functions for each of the four orientation by visual motion direction combinations. These illustrative psychometric functions were fit to the means of all 9 participants. The y-axis represents the percent of rightward or upward responses, while the x-axis represents the contrast ratio of the upward/downward or leftward/rightward visual stimulus (depending on condition). The three curves correspond to each of the three sound conditions (solid = descending, dotted = noise, dashed = ascending). Error bars show standard error between participants. Note that for the on-side leftward/rightward plots, leftward visual motion (i.e., where contrast ratio is equal to 0) corresponds to gravitational upwards relative to the observer.

Hypothesis tests and effect sizes

To directly test whether sound influenced the PSE values in each of the four experimental conditions, within-subjects t-tests were performed to compare the ascending and descending sound conditions in each orientation by direction of visual motion condition. After subtracting the noise condition PSEs from the two sound condition PSEs, the t-tests revealed significant differences for the on-side leftward/rightward condition with $t(8)=3.35$, $p<0.05$, $d=1.0$, on-side downward/upward condition with $t(8)=-3.69$, $p<0.05$, $d=1.9$, upright downward/upward condition with $t(8)=-3.39$, $p<0.05$, $d=2.1$ and an insignificant effect in the upright leftward/rightward condition $t(8)=0.37$, $p>0.05$, $d=0.1$. The Holm-Bonferroni method was used for family-wise correction (Holm, 1979). Since sound had an influence on participants' response behaviour in all conditions except the upright condition making leftward/rightward judgments, it suggests that ascending

and descending tones had an influence on visual motion perception along *both* the body and gravitational reference frames.

Mean PSE scores and their associated absolute mean difference scores across sound conditions (calculated by subtracting the PSEs between sound conditions, averaging the results across participants, and then assigning positive sign) are given for each orientation by visual motion direction condition in Table 1. I chose to present absolute as opposed to signed values as there is no consistent reference frame across the four tasks and thus signs would make reading the data confusing. Comparing the absolute mean differences between the ascending vs. descending sounds across experimental conditions shows that the on-side upward/downward condition had the largest overall effect of sound, followed by the upright upward/downward by a small margin, followed by the on-side leftward/rightward, and then finally the upright leftward/rightward condition.

The absolute mean difference scores comparing the ascending and descending sound conditions from the on-side upward/downward and on-side leftward/rightward conditions will be used as effect sizes in the discussion section as part of a model (see section 3.4). These scores represent how much the descending and ascending sounds biased participants' visual interpretations and pulled their PSEs apart along the individual gravity and body axes. The effect sizes from the on-side leftward/rightward and on-side upward/downward conditions were also statistically compared. The raw PSE difference values from each participant between the ascending and descending sound conditions were thus put into a within-subjects t-test which revealed a significant difference with

t(8)=-3.99, p<0.01. The effect of sounds along the body axis is revealed to be greater than along the gravity axis.

	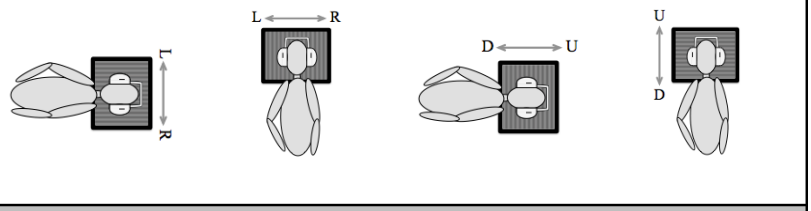			
Mean PSEs				
Asc	0.53	0.50	0.27	0.32
Nz	0.49	0.49	0.51	0.52
Desc	0.46	0.50	0.66	0.69
Absolute mean PSE differences				
Nz vs. Asc	0.04	0.02	0.24	0.20
Asc vs. Desc	0.07	0.005	0.39	0.37
Nz vs. Desc	0.03	0.01	0.15	0.17

Table 1: Mean PSEs for each sound condition (noise, ascending tone, and descending tone) and absolute mean differences in PSE scores between sound conditions for the different orientation and visual motion direction combinations (signified by the cartoons above). Note that the PSEs are visual motion contrast values where 0 represents completely leftward or downward motion, and 1 represents completely rightward or upward motion (depending on condition). Mean PSEs and mean differences calculated and rounded independently of each other.

Response bias control trials

To confirm that the effect of sound biased the visual percept and was not simply due to response bias, the effects of sound in the synchronous and asynchronous trials

were compared. The raw response data (i.e., number of times participants responded upwards or rightwards out of the 20 trials) for the condition where participants sat upright and made upward/downward judgments were used. The three most ambiguous contrast ratios (0.4/0.6, 0.5/0.5, 0.6/0.4) were used (see Figure 15). To compare the effect of sound conditions (ascending tones, descending tones, and noise), the values from the three contrast ratios were pooled and within-subjects t-tests were performed to compare the groups. The tests showed no significant differences between the sound conditions during the asynchronous trials with ascending vs. descending $t(26)=2.62$, $p=0.18$, descending vs. noise $t(26)=1.37$, $p=0.18$, and ascending vs. noise $t(26)= 0.24$, $p=0.80$. In contrast, the comparisons made with the synchronous raw data were all significant with ascending vs. descending $t(26)=9.18$, $p<0.001$, descending vs. noise $t(26)=4.73$, $p<0.001$, and ascending vs. noise $t(26)=5.03$, $p<0.001$. No family-wise multiple comparison corrections were performed on these data so as to show that even under the most liberal conditions the asynchronous comparisons were insignificant. However, with Bonferroni correction over all six conditions (Aickin & Gensler, 1996), all of the synchronous condition comparisons remained significant with $p<0.01$.

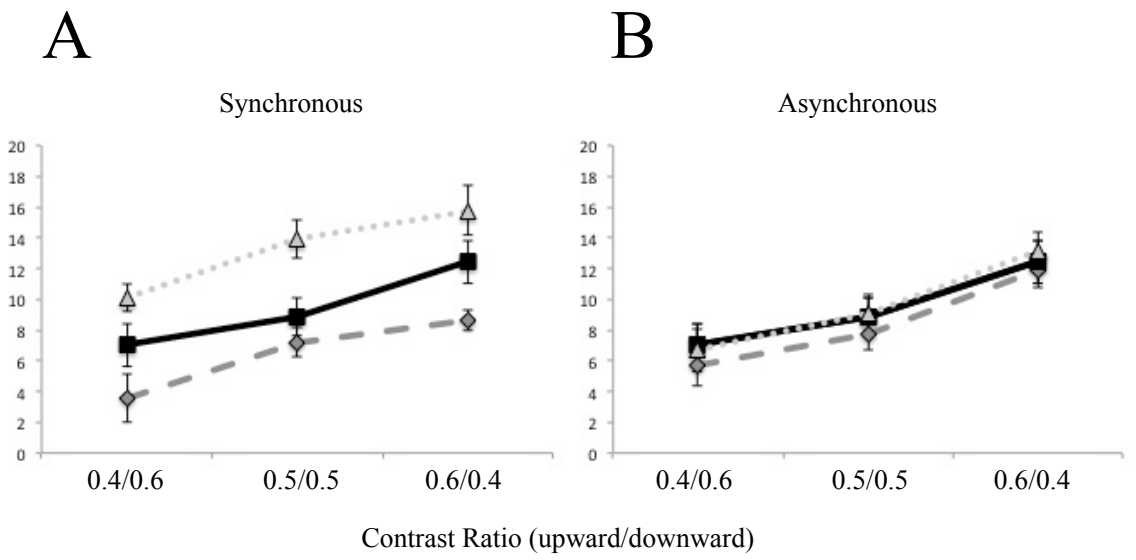

Figure 15. Raw response data in the synchronous (A) and asynchronous (B) conditions across participants, taken from the sitting upright upward/downward visual motion direction condition, with standard error bars. The number of trials where participants responded "upward" are plotted for the three most ambiguous contrast ratios from the ascending (triangle with dotted line), noise (squares with straight line), and descending (diamonds with dashed line) sound conditions.

3.4 Discussion

Significant effects of sound were found in all conditions except the upright condition making leftward/rightward judgments. Since significant effects were found when visual motion was in line with the decoupled body axis (on-side upward/downward visual motion direction) and gravitational axis (on-side leftward/rightward visual motion

direction), the results suggest that binding auditory frequency to spatial elevation for visual motion direction occurs along *both* these axes.

Effects of sounds along the body and gravitational reference frames

Comparing effect sizes for the influence of sounds across the different orientation by visual motion direction combinations suggests that the biasing effect of sound is stronger along the body axis than along the gravitational axis. The magnitude of separation between the ascending and descending sound conditions for the averaged PSEs shows values of 0.07 for the gravitational axis (on-side leftward/rightward judgments) and 0.39 for the body axis (on-side upward/downward judgments, see Table 1). Since both reference frames appear to be used for binding the auditory and visual stimuli, we can hypothesize that the influences of both reference frames are combined in a weighted fashion. This is in line with Parise et al. (2014) as they reasoned that coupling priors (see section 1.4 *statistical correspondences* for explanation of coupling priors) from both the body- and world-centered reference frames were combined in their auditory localization task (see Figure 5B). This is also in agreement with the findings of Roffler and Butler (1968) who found that auditory mislocalization was more common along the head axis than the gravitational axis. The data of experiment 2 are unclear as to whether the hypothesis that the reference frame priors are combined in a weighted fashion is indeed the case. The effect size in the upright upward/downward judgment condition (where both reference frames are aligned) has a nearly identical effect size (0.37), as the body-axis (0.39), under the specific analysis above. It would be expected that when the reference frames are aligned (as in the upright upward/downward task) the effect would be summed and therefore show the greatest effect. If there is such an effect

it is not necessarily linearly summed in the present results, and there may be other factors involved, as the effect sizes of the decoupled body (0.39 along the body axis) and gravitational (0.07) reference frames adds up to 0.46, which is much larger than the actual effect found of 0.36 in the upright upward/downward condition. Regardless, the significant influences of sounds along both the gravitational and body axes suggest that there are priors along both reference frames, similarly to Parise et al. (2014), and the above data are open to interpretation.

If both reference frames are combined in a weighted fashion, than the vector sum model can be used to determine the representation of "upright" the brain is using to combine auditory pitch and visuospatial height. Figure 16A shows the results of the vector sum model based on the current results, where the length of the body and gravity vectors correspond to the effect sizes along these two reference frames. In this case, the model predicts that the multisensory coupling prior along the combined weighted reference frame is 10.2º tilted towards the direction of gravity from the body axis. It is predicted that if the axis of visual motion direction were along this weighted reference frame axis representing "upright", multisensory integration would be optimal, and lead to the largest possible effect. In Maeda et al. (2004) they found that as the direction of visual motion was tilted further and further from spatial upright (corresponding in their case to an aligned body and gravity axis), the effect of sound diminished (Figure 16B). I predict a similar pattern would emerge in reference to the optimal coupling prior reference frame shown in Figure 16A. I suggest that the spatial representation of "upright" used for integrating auditory pitch to visuospatial height is a weighted combination of the body and gravitocentric reference frames.

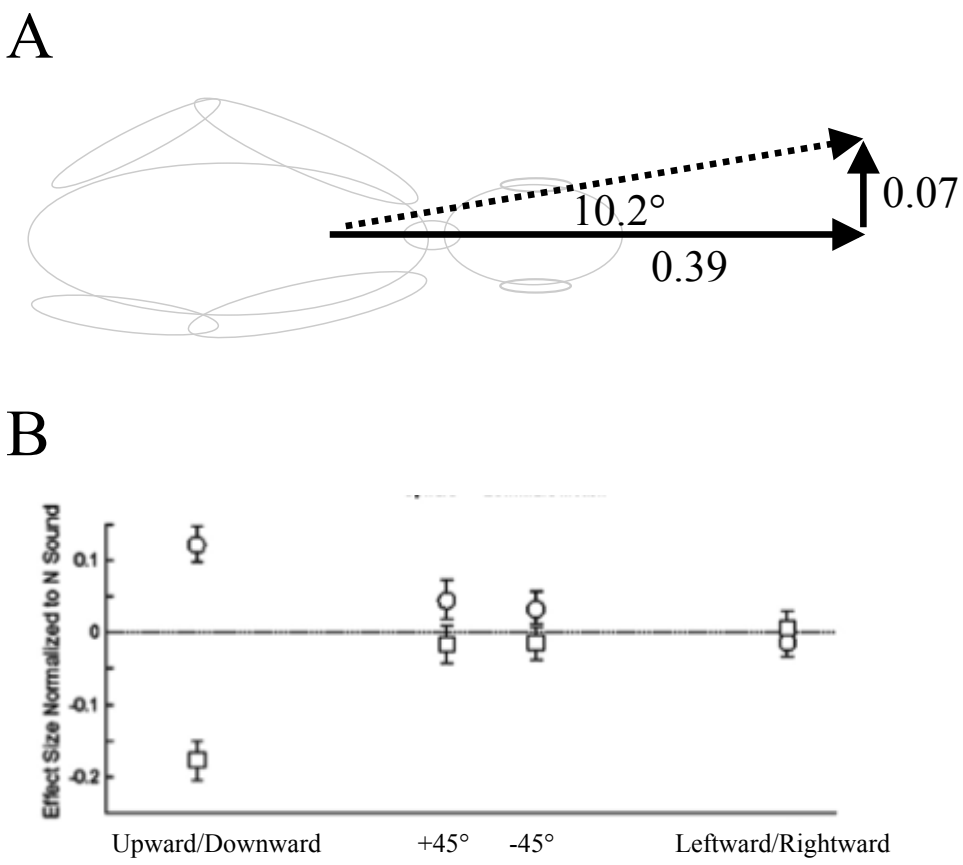

Figure 16. (A) Vector sum model using the effect sizes related to the decoupled body and gravitational reference frames found in the present study. The angle of optimal auditory pitch and visual motion integration is calculated here to be 10.2° towards gravity from the body axis. (B) Data adapted from Maeda et al. (2004) showing normalized effect size data (noise condition PSEs subtracted from the ascending and descending sound condition PSEs) when participants sat upright and the direction of visual motion was varied.

No response bias

Based on comparing the results of the synchronous and asynchronous conditions, it appears that the findings reflect true perceptual effects rather than response bias. Figure 15 shows a clear difference between the synchronous and asynchronous experiments, with clear separation across sound conditions when sounds overlapped temporally with the visual motion compared to when the sounds were played after the visual motion. These results are congruent with Maeda et al. (2004) as they also tested for a range of temporally asynchronous audiovisual stimulus pairings and found no effect when the auditory and visual stimuli did not overlap. Furthermore, Figure 16B, adapted from Maeda et al. (2004), which shows the effect of sound as the orientation of the visual stimulus was tilted away from upright also suggests that the effect is perceptual in nature. In their study, participants had to respond "upward" or "downward" as the orientation of the visual stimulus tilted, and if the effect is based on response bias, than the tilting of the visual stimulus should not have made a difference. Maeda et al. also tested for a semantic effect by presenting the words "Up" and "Down" with the visual stimulus and found no effect.

Possible limitations

There are certain issues in the implementation of experiment 2 that should be addressed. The main concern is the quality of the presentation of the visual stimulus. Based on the distance of the participant's head from the screen and the resolution of the display, it is possible that the visual motion stimulus was not optimal. While the participants' results from this study show that they were able to sufficiently perform the task, the sub-optimal resolution of the display could have led to an image lacking the requisite grain to properly represent the blending of superimposed sinusoid gratings.

While the stimulus was clear enough that participants could perform the task sufficiently, an optimal visual stimulus could have perhaps given more precise results. To improve the visual grain the distance of the participants from the screen could have been increased and the visual gratings respectively adjusted to increase the number of pixels per degree visual angle.

It is worth noting that while this study had nine reliable participants, an equal number could not properly complete the practice trials and were thus not admitted to perform the experimental task. Even during the most unambiguous contrast ratios, they were unable to correctly identify the direction of the upward/downward moving gratings. Interestingly, these nine participants that could not properly complete the practice test with upward/downward visual motion were still able to do the leftward/rightward task adequately. This may suggest an asymmetry in sensitivity between horizontal and vertical motion, or perhaps a difference in sensitivity to spatial frequency at different orientations.

Summary

In summary, since significant effects were found when visual motion was in line with the decoupled body axis (on-side upward/downward visual motion direction) and the gravitational axis (on-side leftward/rightward visual motion direction), the results suggest that binding auditory frequency to visuospatial height for visual motion direction occurs along both these axes. I suggest that each reference frame has its own audiovisual coupling prior, and that these would be combined in a weighted fashion. By using a vector sum model, we can estimate the axis that represents the weighted combination of these two reference frames along which the coupling of auditory pitch and visuospatial

height should be at its maximum. Comparing the synchronous and asynchronous trials shows that the effect is perceptual in nature rather than based on a response bias. Finally, improvements to the implementation of the visual stimulus could potentially make this a more reliable and precise experiment.

4. GENERAL DISCUSSION

4.1 What do the results of experiment 1 and 2 mean, and how do they relate?

The results of experiment 1 (when using either the static- or dynamic-sound stimuli) showed no effect of sounds on the PUs, suggesting that the brain does not use the crossmodal correspondence between auditory pitch and spatial elevation as a cue for perceived self-orientation. In contrast, in experiment 2 the ascending and descending tones indeed biased ambiguous visual motion along both the body and gravitational reference frames (more so along the body axis), suggesting that the crossmodal correspondence between auditory pitch and visuospatial height is integrated with visual motion systematically based on spatial configurations. Figure 16A shows a model of these relative influences where the vectors represent the strength of the pitch-height coupling priors along each reference frame. This model predicts that when the coupling priors from each reference frame are combined in a weighted fashion, there is a visuospatial axis along which auditory pitch and visual motion are optimally combined, and deviations from this spatial orientation gradually lead to weaker audiovisual perceptual integration.

The results from experiment 2 are in-line with the concept of the "mental tonal axis" originally put forth by Rusconi et al. (2006). In their study they showed that when the spatial layout of response keys was incompatible with the mental tonal axis, where higher frequency tones are spatially higher than low tones, participants' speeded judgments of auditory pitch were delayed. Rusconi et al. argued that there was an incompatibility with the spatial mapping. Like most experiments from the literature

exploring how auditory pitch integrates with visual stimuli however, the participants sat upright and the reference frames were not explored.

Here I speculate that the axis of optimal integration that I proposed in section 3.4 is analogous to the mental tonal axis put forth by Rusconi et al. (2006). I predict that when visual stimuli that have a spatial elevation component (i.e., the direction of visual motion, spatial location, etc.) are oriented along the mental tonal axis, the potential for maximum perceptual integration with auditory stimuli varying in frequency can be achieved. I suggest that this will hold true regardless of the specifics of the task, and the effects found in the few experiments that did introduce spatial variations lend weight to this notion, as the experimental paradigms were different (Rusconi et al. 2006; Parise et al., 2014; Maeda et al., 2004; Roffler & Butler, 1968).

Figure 17 shows how the vector sum model from Figure 16A can be generalized to determine the mental tonal axis under a variety of different spatial orientations. With the obtained estimates for the relative weightings of both the body and gravitocentric audiovisual coupling priors, the spatial orientation of the mental tonal axis relative to the observer can be determined. This model predicts that the strength of audiovisual coupling along the mental tonal axis depends on the alignment between the body and gravity vectors. When the reference frames for each of the coupling priors are aligned, the effect should be strongest compared to when they are in conflict to each other. This model predicts that if the axis connecting “high” and “low” visual stimuli is orthogonal to the mental tonal axis, there will be no pitch-height correspondence at all (as in the null effects in the upright leftward/rightward motion condition). Thus, gradually deviating the axis of visual stimuli relative to the mental tonal axis should, as in Maeda et al. (2004),

lead to a gradual decline in the strength of multisensory integration. This is speculation of course, and the results from experiment 2 are unclear as to how the coupling priors may interact. Regardless, a model to predict the orientation of the mental tonal axis (which I also propose is systematically relevant to crossmodal correspondences between pitch and spatial elevation in general) under different spatial configurations is an idea novel to this thesis and the crude details can be worked out in future research.

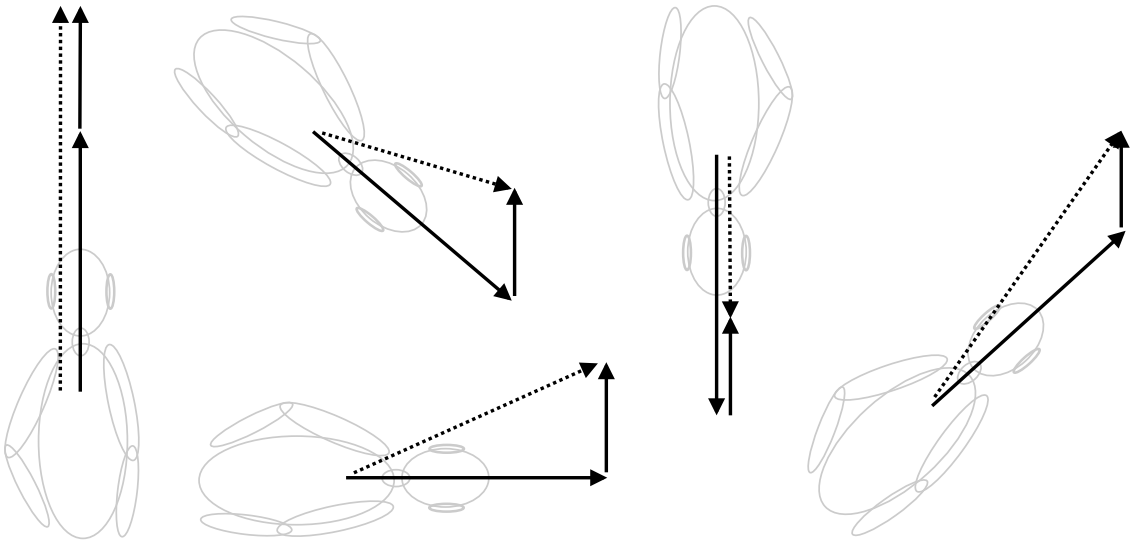

Figure 17. Orientation of the mental tonal axis relative to the observer under different spatial configurations. The solid lines represent the relative influences of the audiovisual coupling priors along both the body and gravitational reference frames. The dotted line shows the orientation of the determined mental tonal axis and the silhouette shows the orientation of the observer. The length of the dotted line predicts the strength of audiovisual coupling under the different conditions where the body and gravitational reference frames are congruent or in conflict with each other.

Here, I propose that the results of experiment 1 and experiment 2 may be related and compatible. The null findings of experiment 1 might be explainable by considering that perhaps the perceptual upright was unaffected by the crossmodal correspondence between pitch and height because the mental tonal axis lies along the perceptual upright. In Dyde et al. (2006), the mean PU while participants laid on their right side was 17° towards gravity, which is essentially consistent with the findings of experiment 1 (with an average mean of 27.3° towards gravity across all four PUs collected). This is roughly consistent with the results of experiment 2, where the axis of optimal integration (with value of 10.2°) is towards gravity while laying on the right side. While the sound stimuli from experiment 1 were intended to act as spatial orientation cues along the gravitational vector, perhaps the binaural aspect of the cue was not integrated with the crossmodal correspondence to affect perception. Perhaps the frequency content of the high- and low-pitched tones was simply represented along the mental tonal axis rather than integrated with external space. As such, the "up" and "down" frequency components would be defined along the mental tonal axis, which my data suggests is essentially in-line with the perceptual upright. Similarly to the spatial properties of the mental tonal axis that I've proposed, the perceptual upright is defined as the vector sum of the body and gravitational reference frames, where the body has a stronger component on average than the gravitational component (Dyde et al., 2006). If they lie along the same spatial axis, and the frequency component of sounds is represented along the mental tonal axis, than there should be no effect of "high" and "low" tones on the perceptual upright.

4.2 Future Research

Based on the themes and results of this thesis, what are possible avenues for future research? First I will mention research ideas that relate directly to the present findings and are intended to clarify and further investigate the present issues. Next I introduce some research ideas that extend beyond what has been discussed thus far in this thesis but are still directly related.

4.2.1 Properties of the predictive model and the mental tonal axis

The first issue to test is the predictions made by the model of the mental tonal axis presented in sections 3.4 and 4 (see Figures 15A and 16). This model can be used to predict the orientation of the mental tonal axis while participants are positioned in different body orientations relative to gravity. By extending experiment 2 and presenting visual motion along the predicted mental tonal axis, a baseline effect of ascending and descending tones can be determined. I hypothesize that the effect of the sound stimuli would be greatest when the axis of visual motion is along the mental tonal axis but gradually diminish as it deviates from the predicted mental tonal axis (as in Maeda et al., 2004). This would provide evidence that the model properly predicts the axis of the mental tonal axis, or otherwise provide a basis for a new model. As discussed in section 3.4, there were certain problems with the setup of experiment 2 that would need to be remedied in order to properly test this prediction. One of the issues was that based on the pre-experimental training period it appeared that participants were more sensitive to leftward/rightward motion compared to upward/downward motion. To properly compare effect sizes for different visual motion orientations, sensitivity to each direction of visual motion would need to be factored into the analysis or dealt with beforehand.

The model from Figure 17 also predicts that as the gravitational and body reference frames misalign and become less congruent with each other, the strength of perceptual coupling between auditory pitch and visual motion presented along the mental tonal axis will decrease (as represented by the length of the dashed line). According to the model, when participants are positioned upside-down and presented with visual motion along the mental tonal axis (here positioned in line with the body reference frame), the effect should be the most diminished as the body and gravitational frames are directly in conflict. I hypothesize that when participants are upside down, sounds will bias perceived motion direction along the body axis, but the effect will be much smaller as the body and gravitational axes are directly in conflict with each other. By comparing the effects of the sound stimuli, while the visual stimulus is always presented along the hypothesized mental tonal axis, under different body orientations, this prediction of the model can be verified.

4.2.2 Extending to 3-D

Related to the spatial characteristics of the mental tonal axis is the issue of whether or not the pitch-elevation correspondence plays a role in perceiving visual stimuli in depth. To put it in broader terms, is the mental tonal axis relevant to perception in three-dimensions? For example, if one is lying supine facing the sky are ascending and descending tones perceived as moving upward and downward only with reference to the body, or also with reference to gravity, which is now represented as towards the sky or closer to the observer? To test this I suggest adapting the visual stimuli of Maeda et al. (2004) and experiment 2 to present optic-flow (Lee, 1980). Instead of presenting upward vs. downward (in reference to the body) visual gratings at varying contrast ratios, use

instead a visual stimulus composed of a mix of optic-flow towards or away at different contrast ratios. Using the same paradigm as experiment 2 except with participants laying on their back completing this task, it can be determined whether or not ascending and descending tones bias the interpretation of towards or away optic-flow when it is aligned with gravity. I hypothesize that the mental tonal axis extends to 3 dimensions and the ascending and descending tones will bias perception of the ambiguous optic flow visual motion stimulus in the directions of away-from and towards the participant, respectively.

4.2.3 Generalizing to different tasks

Another question that is not entirely clear is whether or not the effects of experiment 2 would generalize to different tasks. It has been shown that both auditory localization (Parise et al., 2014) and the interpretation of visual motion are affected by the pitch-height crossmodal correspondence along both the body and gravity reference frames. Also, note that the correspondence in experiment 2 is somewhat distinct in that the association was between ascending and descending (i.e. changing frequency) tones with “upward” or “downward” visual motion. It is uncertain whether or not the spatial constraints discussed in this thesis in relation to the pitch-height association would generalize to other types of tasks. For example, would the effects on the speeded-classification task (Ben-Artzi, 1998) hold up if the visual stimuli were presented “high” and “low” along the gravitational reference frame? The task could be done similarly to experiment 2 with participants tilted on their right side and measure reaction times to the visual target stimuli presented either high or low along both the gravitational and body axes. I hypothesize that all tasks will show effects of using the same mental tonal axis

(consistent with this thesis) for spatially combining auditory pitch and visuospatial height as in experiment 2, and would suggest the use of the same underlying mechanism.

4.2.4 Visual cues to upright and the pitch-height correspondence

The role of visual cues to upright could also be tested using the speeded-classification task. If a presented visual background altered the perception of upright, would the effect of sound on the speeded-classification test be modified? For example, if participants laid on their right side akin to experiment 2 and performed the speeded-classification task with visual stimuli along both the body and gravitational axes, would a visual background with cues to upright along each of these axes bias participants' speeded judgments? I hypothesize that along with both the body and gravitational reference frames playing a role in the pitch-height correspondence, visual cues will also influence the degree to which the auditory and visual stimuli will be perceptually bound. This would also add further evidence that the mental tonal axis lies along the PU, as it is also influenced by all three spatial cues.

4.2.5 Do binaural cues integrate with the pitch-height spatial elevation mapping?

Based on the results of experiment 1, which showed no significant effects of sound on the perceptual upright, it is unclear whether the pitch-height correspondence and binaural cues are used together in perception. Parise et al. (2014) suggested that since the pitch-height correspondence seems to reflect the physical properties of the environment, the external ear might have evolved to accentuate these physical properties for sharper perception. They discussed the spectral filtering properties of the external ear but did not discuss binaural cues. When upright, binaural cues do not carry any

information related to spatial elevation but when laying on one's right side for example, the left ear before the right will sense sounds from above (which should be higher-pitched than sounds that came from below based on the frequency elevation mapping). Has the perceptual system thus adapted to this fact such that binaural cues are taken into account with the pitch-height correspondence? Experiment 1 did not directly test this hypothesis but rather tested whether or not the presented tones would influence perceived-self orientation. My current question is more related to what the brain expects from the environment in terms of binaural as opposed to spectral cues related to the pitch-height correspondence. To test this I propose a speeded-classification task with no visual component. With a setup akin to experiment 1, would randomly presenting high vs. low tones to the left or right ears lead to different reaction times when judging which ear the tone was presented to? I predict increased reaction times when the relative pitch of presented sounds is incongruent with the spatial aspect of ear position (e.g., if the low-pitched sound comes from above). I thus hypothesize that binaural cues are integrated with the reference frames used in the pitch-height correspondence.

4.2.6 Can sounds enhance visual cues to upright?

Also somewhat related to experiment 1 is the question of whether or not sounds could enhance visual cues for perceived orientation. In Dyde et al. (2011), the relative effectiveness of dynamic and static images was compared. They found that as a visual cue for perceiving upright (i.e., significantly influencing the perceptual upright during the OCHART task), video was significantly more effective than a static background image. Both visual stimuli were similar in orientation content but the video which showed people walking around (or standing in the image) led to a stronger effect. Is it possible

that a soundtrack accompanying the video playback would enhance the effectiveness of the video in influencing a participant's perceptual upright? I propose extending the study by Dyde et al. (2001) and compare video with and without accompanying audio. I hypothesize that adding a corresponding auditory soundtrack to a movie with visual cues to upright would significantly enhance the effect of the visual cues.

4.2.7 Can high and low visual stimuli bias the perceived pitch of ambiguous auditory stimuli?

Finally, the essential question of my last proposed research avenue is whether or not the pitch-height correspondence can be demonstrated with an auditory-based task rather than only visual tasks, as have been shown thus far. In other words, is it only the case that presented low- and high-pitched sounds bias visual task performance, or can a spatially low or high visual stimulus bias the interpretation of the pitch of an auditory stimulus? To test this, I suggest presenting an auditory stimulus with an ambiguous pitch, and see if simultaneously presented high or low visual stimuli will bias the auditory perception of pitch. Such an ambiguous auditory stimulus exists and is the subject of research on the tri-tone paradox developed by Deutsche (1987). The tritone paradox is the auditory illusion where a pair of Shepard tones (i.e., a sound consisting of a superposition of sine waves separated by octaves, Shepard, 1964), presented sequentially, and separated by the pitch interval of a tritone (or half-octave in musical terms), are perceived as either ascending or descending in pitch depending on the individual observer's perception, but is in actuality ambiguous. Like the Necker cube (Necker, 1832), the sequentially presented Shepard tone pair is bistable, and can only be perceived as either ascending or descending in pitch, despite the fact that the two Shepard tones are matched in overall

frequency energy. I think this auditory stimulus can be used in a task where a simultaneously presented spatially high or low visual stimulus may bias auditory perception. Basically, present the sequential Shepard tones in tandem with an apparent motion visual stimulus that travels up or down, and ask participants to determine the relative pitch of the second sound. I think this study would be an interesting demonstration that crossmodal correspondences can be bidirectional in nature, where each modality can influence perception in the other. I hypothesize that spatially high and low visual stimuli bias the perception of auditory stimuli that are ambiguous in frequency, and thus the pitch-height correspondence is perceptually bi-directional.

4.3 Applications and relevance of this research

There is value in knowing about crossmodal correspondences in general, both in the realms of commercial and industrial design, as well as in the arts. Here I give a brief overview of how research on crossmodal correspondences can be applied. Specific applications of the pitch-height correspondence are mentioned as well as how the research present in this thesis can be used to benefit these areas.

4.3.1 Commercial design and user-experience

In Don Norman's recent book "The Design of Everyday Things" (2013) he brings together ideas from the cognitive sciences and applies them to the realm of design. He takes theory from various realms of psychology such as positive psychology, cognitive psychology, social/personality psychology, psychology of emotion, perceptual neuroscience, etc., and integrates them with the broad field of design as it applies to technology in general (i.e., both physical and virtual). Norman argues that the user is

often blamed for misuse of technology when in fact the blame should be put on poor design. He believes that improved design can mitigate user error and the resultant negative consequences to the user and those depending on them. His work is currently recognized as a valuable text in the growing commercial field of *user experience design*. User experience design (or UX design) refers to the study of what constitutes a "pleasurable experience" with technology, and the application of these insights in a commercial setting (Hassenzhal et al., 2010). UX has been applied in interface design in software, web design, in the design of everyday things such as media devices, furniture, appliaces, etc., and in the study of user preferences/behaviour in the commercial market.

I think knowledge of crossmodal correspondences can be valuable to UX design. An understanding of the perceptually intuitive relationships between the senses can be used to create user experiences that avoid perceptual conflict and benefit from the ability to predict the resultant cognitive effects. For example, Picqueras-Fiszman & Spence (2011) conducted a study relevant to market research on the product-packaging colour for potato chips. They found that participants more readily associated the colour red with spicy flavours, whereas the colour green was more associated with tart and salty flavours. In another study, Chiou & Rich (2012) found that high- and low-pitched tones directed attentional resources to either a spatially low or high visual object, and could potentially be used in guiding a user's attention. I believe that UX design could benefit from such examples of crossmodal correspondence and their related cognitive effects in designing seamless and enjoyable user experiences.

4.3.2 Industrial design and ergonomics

UX design is directly related to the overarching field of ergonomics, which is generally defined as the study and application of human-artifact interactions, viewed from the unified perspectives of science, engineering, design, technology, and management of human-compatible systems (Karwowski, 2005). UX design incorporates many of these ideas but puts greater emphasis on the users' pleasure and relates more to a general consumer setting whereas the study of ergonomics has a deeper history and applies most broadly, including industrial or employment-related settings. Research on crossmodal correspondences can be applied here in the same ways as in UX design mentioned above. The pitch-visuospatial height correspondence for example has been studied in the realm of ergonomics research for interface design, where the correct mapping of crossmodal cues is meant to lead to more efficient behavioural outcomes (Rusconi et al., 2006; Dutta & Proctor, 1992). My research is directly applicable here as it reveals some of the spatial properties and constraints of this perceptual association. My research shows that this association depends on both the body reference frame and the gravitational reference frame. Thus, under different body or gravitational states, the orientation of the mental tonal axis may be modified and the intended design of the technology may come undone, leading to unfavourable outcomes (e.g., Endsley & Rosiles, 1995). One such form of technology that could benefit from knowledge on crossmodal correspondences is technology based on research of sensory substitution, technologies where a device is used to present information typically acquired from one sense to another sense, often to substitute for its unavailability (Bach-y-Rita & Kercel, 2003). An example is a prototype by Meijer (1992), which converts video into audio mapping, where visual height is associated with pitch, and brightness with loudness.

Knowledge of crossmodal correspondences thus can aid in developing sensory mappings from one sense to the other that are most intuitive.

4.3.3 Crossmodal correspondences and the arts

Finally, I suggest that crossmodal correspondences can be applied in the arts, and probably already are. For example, in designing a multi-media experience for a music concert, the colours of a lightshow could be used strategically to evoke certain emotional or perceptual states that are meant to enhance the musical experience of the viewer. With the wide range of audiovisual crossmodal correspondences there is opportunity to create some compelling perceptual experiences that may surprise consumers of media. Another example, outside of the audiovisual domain, is natural associations between tastes and visual shape applied to the culinary arts.

4.4 Final remarks

As shown above, crossmodal correspondences have potential for practical application in realms such as design and technology. But perhaps even more importantly in my opinion, crossmodal correspondences are phenomenologically rich. They remind us to pay attention to the detail of our consciously perceived experiences, both appreciating what each sense has to offer individually, and also in illustrating how the senses are fundamentally interconnected. There is beauty in the notion of harmonious relationships between fundamentally different entities and the observation that seemingly arbitrary perceptual juxtapositions can be aesthetically true. This notion can even be extended to the metaphysical, where the tying together of perceptions within the

individual mirrors the natural interrelationships between the physical features of the outside world, and vice-versa.

References

Adams, W.J., Grad, E.W., & Ernst, M.O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience*, 7(*10*), 1057-1058.

Aickin, M. & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. *American Journal of Public Health*, 86(*5*), 726-728.

Algazi, V.R., Duda, R.O, Thompson, D.M., & Avendano, C. (2001). The CIPIC HRTF database. 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, (IEEE, New Paltz, NY), 99-102.

Aubert, H. (1861). Eine scheinbare Drehung von Objekten bei Nei- gung des Kopfes nach rechts oder links. *Virchows Archiven*, 20, 381-39.

Bahrick, L.E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, 13(*3*), 99-102.

Bach-y-Rita, P.B. & Kercel, S.W. (2003). Sensory substitution and the human-machine interface. *TRENDS in Cognitive Science*, 7(*11*), 541-547.

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of $1/f$ noise. *Physical Review Letters*, 59(*4*), 381-384.

Batteau, D.W. (1967). The role of the pinna in human localization. *Proceedings of the Royal Society of London*, B168(*11*), 158-180.

Belkin, K., Martin, R., Kemp, S.E., & Gilbert, A.N. (1997). Auditory pitch as a perceptual analogue to odor quality, *Psychological Science*, 8, 340-342.

Ben-Artzi, E. & Marks, L. E. (1995). Visual–auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, 57, 1151-1162.

Bernstein, I.H. & Edelstein, B.A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, 87(*2*), 241-247.

Bertelson, P., Vroomen, J., Wiegeraad, G., & de Gelder, B. (1994). Exploring the relation between McGurk interference and ventriloquism. *Proceedings of the 1994 International Conference on Spoken Language Processing*, 2, 559-562.

Blauert, J. (1974). *Spatial Hearing: Revised Edition, Massachusetts Institute of Technology*, United States of America.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436.

Bremner, A.J., Caparos, S., Davidoff, J., Fockert, J., Linnell, K.J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, 126(*2*), 165-172.

Bronner, K. (2011). *What is the sound of citrus? Research on the correspondences between the perception of sound and taste/flavor.*

Bronner, K., Bruhn, H., Hirt, R., & Piper, D. (2008). *Research on the interaction between the perception of music and flavor*. Poster presented at the 9th Annual Meeting of the International Multisensory Research Forum (IMRF), Hamburg, German, July.

Burge, J., Geisler, W.S. (2011). Optimal defocus estimation in individual natural images. *Proceedings in the National Academy of Sciences*, 108(*40*), 16849-16854.

Chen, Y.C. & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, 114(*3*), 389-404.

Chiou, R. & Rich, A.N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, 41(*3*), 339-353.

Crisinel, A.S. & Spence, C. (2009). Implicit association between basic tastes and pitch. *Neuroscience Letters*, 464, 39-42.

Crisinel, A.S. & Spence, C. (2009). A fruity note: Crossmodal associations between odors and musical notes. *Chemical Senses*, 1-8, doi:10.1093/chemse/bjr085

Davis, R. (1961). The fitness of names to drawings: A cross-cultural study in Tanganyika. *British Journal of Psychology*, 52(*3*), 259-268.

Deutsch, D. (1987). The triton paradox: Effects of spectral variables. *Perception & Psychophysics*, 41(*6*), 563-575.

Doehrmann, O. & Naumer, M.J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain research*, 1242(*25*), 136-150.

Dutta, A. & Proctor, R. W. (1992). Persistence of stimulus-response compatibility effects with extended practice. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 801–809.

Dyde, R.T., Jenkin, M.R., & Harris, L. (2006). The subjective visual vertical and the perceptual upright. *Experimental Brain Research*, 173, 612-622.

Dyde, R.T, Jenkin, H.L., Zacher, J.E., & Harris, L.R. (2011). Perceptual upright: The relative effectiveness of dynamic and static images under different gravity states. *Seeing and Perceiving*, 24, 53-64.

Easton, D.R., Greene, A.J., DiZio, P., & Lackner, J.R. (1998). Auditory cues for orientation and postural control in sighted and congenitally blind people. *Experimental Brain Research*, 118, 541–550.

Eitan, Z. & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, 114(*3*), 405-422.

Eitan, Z., Schupak, A., & Marks, L.E. (2008). Louder is higher: Cross-modal interaction of loudness change and vertical motion in speeded classification. In K. Miyazaki, Y.

Hiraga, M. Adachi, Y. Nakajima, & M. Tsuzaki (Eds.), Proceedings of the 10th international conference on music perception and cognition (ICMP10). Adelaide: Causal Productions

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.

Endsley, M.R. & Rosiles, A. (1995). Auditory localization for spatial orientation. *Journal of Vestibular Research*, 5(*6*), 473-485.

Ernst, M.O. (2006). A Bayesian view on multimodal cue integration. In G. Knoblich, I.M. Thornton, M. Grosjean, & M. Shiffrar (Eds.) *Human body perception from the inside out* (pp. 105-131). Oxford: Oxford University Press.

Ernst, M.O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7(*5*), 1-14.

Ernst, M.O. & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(*6870*), 429-233.

Ernst, M.O. & Bülthoff, H.H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(*4*), 162-169.

Evans, K. K. & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, 10(*1*), 6:1–12.

Fesselier, R.Y. & Knoblauch, K. (2006). Modeling psychometric functions in R. *Behaviour Research Methods*, 38(*1*), 28-41.

Field, A., Miles, J., & Field, Z. Discovering statistics using R. (California: Sage Publications, 2012), 511-530.

Gallace, A. & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of viual size. *Perception & Psychophysics*, 68(*7*), 1191-1203.

Garner, W.R. (1974). *The processing of information and structure*. Potomac, MD: Laurence Elrbaum Associates.

Gebels, G. (1969). An investigation of phonetic symbolism in different cultures. Journal of Verbal Learning and Verbal Behaviour, 8, 310-312.

Gilbert, A.N., Martin, R. & Kemp, S.E. (1996). Cross-modal correspondence between vision and olfaction: The color of smells. *The American Journal of Psychology*, 109, 335-351.

Green, A.M. & Angelaki, D.E. (2010). Multisensory integration: Resolving sensory ambiguities to build novel representations. *Current Opinion in Neurobiology*, 20(*3*), 353-360.

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers*, 22(*5*), 353-362.

Holt-Hansen, K. (1968). Taste and pitch. *Perceptual and Motor Skills*, 27, 59-68.

Holt-Hans, K. (1976). Extraordinary experiences during cross-modal perception. *Perceptual and Motor Skills*, 43, 1023-1027.

Howard, I.P. (1982). *Human visual orientation*, Wiley, New York.

Hinton, L., Nichols, J., & Ohala, J.J. (Eds.). (1994). *Sound symbolism*. Cambridge, UK: Cambridge University Press.

Innes-Brown, H. & Crewther, D. (2009). The impact of spatial incongruence on an auditory–visual illusion. *PLoS ONE*, 4, e6450.

Jones, J.A. & Jarick, M. (2006). Multisensory integration of speech signals: The relationship between space and time. *Experimental Brain Research*, 174, 588-594.

Jones, J. A. & Munhall, K. G. (1997). The effects of separating auditory and visual sources on audiovisual integration of speech. *Canadian Acoustics*, 25, 13-19.

Karwowski, W. (2005). Ergonomics and human factors: the paradigms for science, engineering, technology, and management of human compatible systems. *Ergonomics*, 48(*5*), 436-463.

Kemp, S.E., & Gilbert, A.N. (1997). Odor intensity and color lightness are correlated sensory dimensions. *The American Journal of Psychology*, 110, 35-46.

Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.

Lackner, J.R. & DiZio, P. (2005). Vestibular, proprioceptive, and haptic contributions to spatial orientation. *Annual Review of Psychology*, 56, 115-147.

Lee, D.N. (1980). The optic flow field: The foundation of vision. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 290(*1038*), 169-170.

Lewkowicz, D.J. & Turkewitz, G. (1980). Cross-modal equivalency in early infancy: Auditory-visual intensity matching. *Developmental Psychology*, 16, 597-607.

Lewald, J. & Ehrenstein, W.H. (1998). Influence of head-to-trunk position on sound localization. *Experimental Brain Research*, 121, 230-238.

Long, J. (1977). Contextual assimilation and its effect on the division of attention between nonverbal signals. *The Quarterly Journal of Experimental Psychology*, 29(*3*), 397-414.

Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology*, 14, R990–R991.

Marks, L.E. (1978). *The unity of the senses: Interrelations among the modalities*. New York: Academic Press.

Marks, L.E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded classification. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 384-394.

Marks, L.E. (1989). On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness. *Journal of Experimental Psychology: Human Perception and Performance*, 15(*3*), 586-602.

Marks, L.E. (2004). Cross-modal interactions in speeded classification. In G.A. Calvert, C. Spence & B.E. Stein (Eds.), *The handbook of multisensory processes*. (pp. 85-106). Cambridge, MA: MIT Press.

Martino, G. & Marks, L.E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*, 28(*7*), 903-924.

Martino, G. & Marks, L.E. (2000). Cross-modal interaction between vision and touch: The role of synaesthetic correspondence. *Perception*, 29, 745-754.

Melara, R. D. & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116, 323–336.

Morgan, G.A., Goodson, F.E., & Jones, T. (1975). Age differences in the association between felt temperatures and color choices. *The American Journal of Psychology*, 88, 125-130.

Mesz, B., Trevisan, M., & Sigman, M. (2011). The taste of music. *Perception*, 40(*2*), 209-219.

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, 17(*1*), 160-169.

Mittelstaedt, H. (1983). A new solution to the problem of the subjective vertical. *Naturwissenschaften*, 70, 272–281.

Mondloch, C.J. & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in your children. *Cognitive, Affective, and Behavioural Neuroscience*, 4, 133-136.

Myung, I.J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47, 90-100.

Necker, L.A. (1832). Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. *London and Edinburgh Philosophical Magazine and Journal of Science*, 1(*5*), 329-337.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. New York, Basic Books.

Osgood, C.E., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Patching, G. R. & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 755–775.

Parise, C. (2012). *Signal compatibility as a modulatory factor for audiovisual multisensory integration.* (Doctoral Dissertation).

Parise, C.V., Knorre, K., & Ernst, M.O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, 111(*16*):6104-8

Piesse, C.H. (1891). *Piesse's art of perfumery (5th ed.)*. London: Piesse and Lubin.

Picqueras-Fiszman, B. & Spence, C. (2011). Crossmodal correspondences in product packaging. Assessing colour-flavor correspondences for potato chips (crisps). *Appetite*, 57, 753-757.

Pratt, C.C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, 13, 278-285.

Raab D.H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24(*5*), 574-590.

Radeau, M. & Bertelson, P. (1994). Auditory-visual interaction and modularity. *Current Psychology and Cognition*, *13*, 3-51.

Ramachandran, V. & Hubbard, E. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies*, 8(*12*), 3-34.

Robinson, D.W. & Dadson, R.S. (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7, 166-181.

Roffler, S. K. & Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *The Journal of the Acoustical Society of America*, 43, 1255–1259.

Rogers, S.K. & Ross, A.S. (1968). A cross-cultural test of the Maluma-Takete phenomenon. *Perception*, 4(*1*), 105-106.

Rouw, R. & Scholte, H.S. (2007). Increased structural connectivity in grapheme-color synesthesia. *Nature Neuroscience*, 10(*6*), 792-797.

Rudmin, F. & Cappelli, M. (1983). Tone-taste synesthesia: A replication. *Perceptual and Motor Skills*, 56, 118.

Rusconi, E., Kwan, B., Giordano, B.L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, 99, 113-129.

Shepard, R.N. (1964). Circularity in judgements of relative pitch. *Journal of the Acoustical Society of America*, 36, 2345-2353.

Shore, D. I., Barnes, M. E., & Spence, C. (2006). The temporal evolution of the crossmodal congruency effect. *Neuroscience Letters*, 392, 96–100

Simner, J. & Ludwig, V. (2009). What colour does that feel? Cross-modal correspondences from touch to colour. Paper presented at the Third International Conference of Synaesthesia and Art, Granada, Spain, April.

Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mapping across gustation and audition. *Perception*, 39, 553-569.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, 28(*2*), 61-70.

Spence, C., Levitan, C.A., Shankar, M.U., & Zampini, M. (2010). Does food color influence taste and flavor perception in humans? *Chemosensory Perception*, 3, 68-84.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73, 971-995.

Stevens, S.S. (1957). On the psychophysical law. *Psychological Review*, 64(*3*), 153.

Stocker, A.A. & Simoncelli, E.P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(*4*), 578-585.

Stumpf, K. (1883). *Tonpsychologie*. Leipzig: S. Hirzel.

Taylor, I.K. & Taylor, M.M. (1962). Phonetic symbolism in four unrelated languages. *Canadian Journal of Psychology*, 16, 344-356.

Treisman, A. & Gelade, G. (1980), A feature integration theory of attention. *Cognitive Psychology*, 12(*1*), 97-136.

Trimble, O.C. (1934). Localization of sound in the anterior-posterior and vertical dimensions of "auditory" space. *British Journal of Psychology*, 24, 320-335.

Väljamäe, A. (2009). Auditorily-induced illusory self-motion: A review. *Brain Research Reviews*, 61, 240-255.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory–visual speech perception. *Neuropsychologia*, 45, 598–607.

Von Hornbostel, E.M. (1931). Uber Geruchshelligkeit [On odour/smell brightness]. *Pflugers Archiv fur Gesamte Physiologie*, 227, 517-538.

Walker, P. & Smith, S. (1985). Stroop interference based on the multimodal correlates of haptic size and auditory pitch. *Perception*, 14, 729-736.

Walker, P., Bremmer, G.J., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, P.S. (2009). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21(*1*), 21-25.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7(*11*), 483-488.